

DOCUMENT RESUME

ED 436 541

TM 028 166

TITLE The Impact of First Steps on the Reading and Writing Ability of Year 5 Students.

INSTITUTION Australian Council for Educational Research, Victoria.

SPONS AGENCY Western Australia Ministry of Education, Perth. Curriculum Development Branch.

PUB DATE 1993-05-00

NOTE 83p.; For related "First Steps" documents, see TM 028 165-170 and TM 028 176.

PUB TYPE Reports - Evaluative (142)

EDRS PRICE MF01/PC04 Plus Postage.

DESCRIPTORS Child Development; *Developmental Stages; Elementary School Students; *Elementary School Teachers; Foreign Countries; Grade 5; High Risk Students; Intermediate Grades; *Literacy; Program Evaluation; *Reading Achievement; Teacher Attitudes; *Writing (Composition); Writing Tests

IDENTIFIERS Australia (Western Australia); *Continuum Models; *First Steps Program (Australia); Impact Evaluation

ABSTRACT

This is one of a series of reports that document the formative research that supported the development of the "First Steps" program designed to improve the literacy and numeracy of primary school students in Western Australia, particularly those considered at-risk. The four themes of First Steps (reading, writing, spelling, and oral language) are organized around a development continuum of milestones along the child's path to literacy. This interim report presents preliminary findings from two tests administered to First Steps students as the first part of a time series study. The main aim of this report is to examine the impact of First Steps on the reading and writing ability of Year 5 students while taking into account the effect of socioeconomic status background. Reading comprehension data were available from 665 year-5 students in 31 classes in 19 schools, and writing ability data came from 736 students. There was evidence that First Steps is making a difference in the reading ability of students, with suggestions that the effect of First Steps seems stronger for children of low socioeconomic status background. Results from the writing exercise suggest that First Steps is having a positive effect on the writing ability of children regardless of their socioeconomic status. Eight appendixes present statistical information, details about the reading and writing tests, and further descriptions of analyses procedures. (Contains 19 exhibits and 3 references.) (SLD)

The Impact of First Steps on the Reading and Writing Ability of Year 5 Students

Im028166

ACER

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

Dennis Jackson

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Editor's Note

This document is one of a series of reports that document the formative research that supported the creation and development of *First Steps*TM. As a result of this research, the Education Department of Western Australia (EDWA), in collaboration with the Australian Council for Educational Research (ACER) revised *First Steps* in response to each of the issues and questions raised by this research. *First Steps* training courses, Developmental Continua, and Resource Books are published with due amendments and alterations.

Other research documents that support the development of *First Steps* include:

Dr. Phil Deschamp:

- A Survey of the Implementation of the Literacy Component of the *First Steps* Project in WA
- The Implementation of The Literacy Component of The *First Steps* Project in ELAN Schools
- A Survey of the Effectiveness of the Focus Teacher 'B' Training for the *First Steps* Project
- Student Achievement: A Study of the Effects of *First Steps* Teaching on Student Achievement
- Case Studies of The Implementation of the *First Steps* Project in Twelve Schools
- The Development and Implementation of the *First Steps* Project in Western Australia

ACER:

- Empirical Validation of the *First Steps* Reading Continuum
- Empirical Validation of the *First Steps* Spelling and Writing Continua
- Empirical Re-Validation of the *First Steps* Spelling Continuum
- Assessment and Record of the Changes made to the Spelling Continuum
- The Impact of *First Steps* on Schools and Teachers
- The Impact of *First Steps* on the Reading and Writing Ability of Year 5 Students
- Background: *First Steps* and the ACER Evaluation & Report on the Validity of the *First Steps* Writing and Spelling Continua*

EDWA:

- Supporting Linguistic and Cultural Diversity Through *First Steps*: The Highgate Project

For more information about on-going *First Steps* research, please contact:

*First Steps*TM / Heinemann
361 Hanover Street
Portsmouth, NH 03801-3912
1.800.541.2086, ext. 281
firststeps@heinemann.com
www.heinemann.com/firststeps

**THE IMPACT OF FIRST STEPS ON THE READING AND WRITING
ABILITY OF WESTERN AUSTRALIAN YEAR 5 SCHOOL STUDENTS**

an interim report to the
Curriculum Development Branch
Western Australian Ministry of Education

Australian Council for Educational Research, May 1993.

This is the second of three reports on the ACER 1992 evaluation of the First Steps project in Western Australia. The first report describes the empirical validation of the First Steps Spelling and Writing continua. The third report examines the responses of teachers to various aspects of the First Steps project. Each of these reports is available from the Curriculum Development Branch of the Ministry of Education.

Thanks to all the schools who agreed to participate in the evaluation of First Steps, and thanks to all the students, teachers, principals and administrative staff in schools who so generously contributed towards it. In particular thanks to the students who completed the writing and reading tests and the teachers who administered and commented upon them. The results from these tests are used in this report.

SUMMARY OF THE MAIN FINDINGS

The following conclusions can be drawn from the data analysed in this report:

- If students from old First Steps schools had lower average levels of reading comprehension ability than students in new or non First Steps schools before First Steps was introduced, then it is likely that First Steps has lead to an improvement in the reading ability of those students exposed to the program.
- If students from old First Steps schools had lower average levels of writing ability than students in new or non First Steps schools before First Steps was introduced, then it is likely that First Steps has lead to an improvement in the writing ability of those students exposed to the program.
- There is some evidence that First Steps may be making an important difference in the reading ability of students. Some informed guessing of likely mean scores before the introduction of First Steps suggests that the mean TORCH reading score of children in old First Steps schools may have moved from 'the low' range to within in the 'average' range of scores. This improvement seems to be due to the effects of First Steps.

CONTENTS

Introduction

• Background to the First Steps Project	1
• Background to the evaluation of the First Steps Project	1
• Aims of the report	1
• The sample of students	2
• Research Questions	3
• Research design	4
Two issues associated with the research design:	
1. A logical fallacy	4
2. Name for the 'non low' socio-economic status background group	4
• Data collection instruments:	6
The TORCH reading comprehension test:	6
Scoring the TORCH test	6
Selection of the texts used with the TORCH test	7
Teachers' responses to the use of the TORCH test	8
The writing exercise:	9
Scoring the writing exercise	10
Teachers' responses to the use of the writing exercise	11

Data Analysis

• Limitations of the data	
The effect of changed circumstances on how the data are analysed	13
The underlying logic of the data analysis	13
Graphing the main effects and the interaction effects	14
The effect of the changed role of the data on the interpretation of TORCH reading comprehension scores	15
A preliminary note on statistical significance versus practical or theoretical significance	23
• Organisation of the data analysis	24
• Distribution of TORCH reading comprehension scores	24
• Distribution of Writing exercise scores	25

• Analysis of TORCH reading comprehension scores	27
The spread of the mean TORCH scores	29
The apparent interaction effect	29
The order of the mean TORCH scores for low socio-economic status background students.	30
The order of the mean TORCH scores for middle or high socio-economic status background students.	32
A comparison between the mean scores of the low and of the middle or high socio-economic status background students	33
Conclusions based on TORCH data	33
• Analysis of the Writing exercise scores	34
The order of the median writing scores for low socio-economic status background students	34
The order of the median writing scores for middle or high socio-economic status background students	35
Conclusions from writing exercise data	
 Conclusion	 36
 Bibliography	 37

APPENDICES

- Appendix A. Weighting of the sample of Year 5 students. Appendix A is printed on pink paper.
- Appendix B. Texts and answer sheets used with the TORCH reading comprehension test. Instructions to the teachers are also attached. Appendix B is printed on blue paper.
- Appendix C. Graphical display showing the estimated range of difficulty for each of the texts used with the TORCH reading comprehension test. Appendix C is printed on dark green paper.
- Appendix D. Text of the forms used for the Writing exercise, including instructions to the teachers. Appendix D is printed on light green paper.
- Appendix E. The ACER *Criteria for the Assessment of Writing*. Appendix E is printed on buff paper.
- Appendix F. Percentile ranks and stanines compared with standard deviations under the normal curve. Appendix F is printed on orange paper.
- Appendix G. The reason for not using tests of statistical significance. Appendix G. is printed on dark blue paper
- Appendix H. A description of the ANOVA procedure. Appendix H. is printed on dark pink paper

INTRODUCTION

Background to the First Steps Project

First Steps is a program instituted by the WA Ministry of Education to improve the literacy and numeracy of primary school students. It is intended, in particular, to assist in the development of the literacy and numeracy skills of 'at risk' students. First Steps was introduced in 1988 and has been evolving since that time. So far most of the work produced as part of the First Steps project has focussed upon literacy.

Background to the evaluation of First Steps

The WA Ministry of Education approached ACER in early 1992 to evaluate First Steps. Work for the evaluation began in April 1992. The evaluation was designed to measure the effects of First Steps at three levels. These levels were: (1) the school (2) the teachers and (3) the students. Questionnaires were used to gather data to measure the effects of First Steps at the school and teacher level. Reading and writing tests were used to try and establish the effects of First Steps at the student level. It is the data taken from these tests that are reported here. The evaluation was also designed to validate the First Steps continua. (See *The Impact of First Steps on Schools and Teachers* and *Empirical Validation of the First Steps Spelling and Writing Continua*.¹)

Aims of this report

This is an interim report. It aims to provide preliminary findings derived from two tests that were administered as the first part of a time series study. This study aims to examine the impact of First Steps on the reading and writing abilities of children. Originally, it was planned to administer one round of tests each year for three years. At the time of this report being prepared, however, it is unclear whether the next two rounds of tests will be administered. Consequently, the data presented here are explored on the assumption that these are likely to be the only test data measuring reading and writing ability of children available for this evaluation of First Steps. Within these limitations, the main aim of this report is to examine what impact First Steps has on the reading and writing ability of Year 5 students while taking into account the effect of socio-economic status background.

¹These reports are part of the series of three reports produced as part of the ACER 1992 evaluation of First Steps.

The sample of students

The sample was stratified into six groups based on the level of disadvantage of the schools and the level of involvement of the school in First Steps. Exhibit 1 shows how the six strata were derived.

Exhibit 1: Strata used in drawing the sample of Year 5 students for the administration of reading and writing tests.

Disadvantage status of school	First Steps status of school		
	Non First Steps	New First Steps	Old First Steps
Disadvantaged (PSP)	Non FS PSP	New FS PSP	Old FS PSP
Non Disadvantaged	Non FS non PSP	New FS non PSP	Old FS non PSP

A non First Steps school was defined as one that had had no formal involvement with First Steps. It should be noted, however, that within these non-First Steps schools there were many teachers who used First Steps materials and ideas.² A new First Steps school was defined as one with less than 12 months formal involvement with First Steps. The level of disadvantage of a school was defined using the Ministry of Education's formal definition of disadvantage - its 'PSP' status.

The schools in each stratum were selected with a probability proportional to size. In other words, the more students there were in a school, the more likely the school was to be chosen. If school size is not taken into account when sampling then small schools tend to be over represented in the sample. Four schools were selected from each stratum, so a total of 24 schools was chosen.

Unless there were more than two Year 5 classes in a school all Year 5 students within a school were included in the sample. Where there were more than two Year 5 classes, two classes were randomly selected and only students from these two classes were chosen. A Year 5 class was defined as any class in which there was at least one Year 5 student. Thus, Year 5 students from composite classes were sometimes selected.

Not all selected schools or classes within schools chose to participate in the evaluation. The final set of reading comprehension data consisted of 665 valid responses from 31 classes in 19 schools. The writing ability data consisted of 736 valid responses from 33 classes in 19

²Responses to questionnaires sent to teachers within these schools showed that nearly two thirds of Year 1 and Year 5 teachers used some First Steps ideas. Some had recently come from schools that had implemented First Steps and others had obtained First Steps materials from colleagues and used these materials without having had any formal involvement with First Steps.

schools. One school that participated in the TORCH testing did not participate in the writing exercise and one that participated in the writing exercise did not participate in the TORCH testing. It was not possible to establish precisely how many of the same classes participated in both the TORCH test and the writing exercise³ although, as nearly all schools in the sample only had one or two Year 5 classes, it is likely that nearly all classes completing the TORCH test also completed the writing exercise.

To generalise⁴ to the population of Year 5 government school students in Western Australia⁵ the achieved samples were weighted so that the numbers in each stratum reflected the proportions of students in each stratum of the population. Details of the weighting procedures are provided in Appendix A.

Research questions

Because of limitations in the data (they are only the first set of a proposed time series study) an exploratory approach to the data is taken, somewhat akin to a case study approach where complexities and issues are highlighted rather than questions answered. Such an approach will lead to quite tentative results. Within this set of constraints, the aim of this report is to address the following questions:

Is there a difference between *the reading comprehension skills* of students in non First Steps, new First Steps and old First Steps schools while taking into account the effect of socio-economic status of the students?

Is there a difference between *the writing skills* of students in non First Steps, new First Steps and old First Steps schools while taking into account the effect of socio-economic status of the students?

If differences are found between these groups of students, what other evidence is available that will assist in making judgements about the importance of these differences?

³This was because some teachers wished to remain anonymous.

⁴Strictly, it is unnecessary to weight the data used in this report because no generalisable claims based on statistical testing are made. The data are weighted because if other data are collected and the set formed into a time series study, then this complete set of data will be weighted. It is important that all published results from the same set of data are based on consistent procedures.

⁵More precisely, the population consists of all Year 5 students in WA government primary and District High schools that are not in the Kimberley region. Kimberley schools were sampled separately for another part of the evaluation. They were not asked to provide test data. They were excluded from the population here so that there was consistency in the definition of populations for each of the various parts of the evaluation of First Steps.

Research design

The planned research design is a time series study examining the reading and writing ability of Year 5 students in Western Australian school students.

This interim report examines the data collected from the first of three planned rounds of testing. Consequently, just a snapshot of the reading and writing ability of students is obtained.

Non First Steps schools were selected so that the achievements of students in First Steps schools could be compared with students at schools not involved in the program.

New First Steps and old First Steps schools were selected in an attempt to capture how the program modifies the writing and reading ability of students over time. By choosing these schools it was hoped that by the end of the planned three year cycle of testing a picture could be built up of the impact of First Steps in the first three years at a school (new First Steps schools) and for longer periods in the old First Steps schools. It was hoped to be able to trace the impact after six years in the oldest First Steps schools. These changes would then be compared with changes in non First Steps schools.

Two issues associated with the research design

A logical fallacy

The PSP status of schools was used to estimate the socio-economic status background of the students. The socio-economic status background of students is known to be related to levels of educational achievement. It is important, therefore, to take into account the effect of socio-economic background when examining the First Steps program.

Using the PSP status of a school is a practical and convenient measure of the socio-economic status background of students. However, its use also presents a problem. To ascribe a school's level of disadvantage to the children within that school is to commit a logical fallacy. This fallacy occurs when the attributes of an environment or a group are ascribed to an individual within that environment or group.⁶ One way of dealing with this problem is to treat the

⁶An example of this fallacy might make it clearer. Consider the following argument:

X lives in suburb Y
Y is a wealthy suburb
Therefore X is wealthy.

The claim that X is wealthy is invalid. (X may well be wealthy but this cannot be established by using this invalid argument.)

relationship between the school's attributes and the student in terms of probabilities. Thus, if it is assumed that the children in disadvantaged schools are *more likely* to have a low socio-economic background then it is not being claimed that any one particular child has the socio-economic status background of the school. Instead the weaker and more defensible claim is being made that, on average, children in disadvantaged schools will have a low socio-economic background.⁷

When the data are collected together - as they are for all the analyses that follow - it is assumed that, on average, children from a disadvantaged school will have a lower socio-economic status than children, on average, from non disadvantaged schools.

The name for the 'non low' socio-economic status background group

Throughout the rest of the report students are referred to as coming from a 'low socio-economic status background' when they attend a disadvantaged (PSP) school. All other students are regarded as coming from a 'non low' economic status background. It was felt, however, that to describe this group as 'non low' was clumsy. Another option, to call this group 'high socio-economic background' (the opposite of 'low') was judged to be misleading since this implies that the middle ranges are excluded. Consequently, it was decided to call the 'non low' group the 'middle or high socio-economic status background' group.

⁷Following from the example in the preceding footnote, the form of the argument now becomes:

If X lives in suburb Y, and,
Y is a wealthy suburb, and
X is a randomly drawn sample of individuals, and

If, A lives in suburb B, and,
B is a poor suburb, and
A is a randomly drawn sample of individuals

then on average individuals in sample X will be more wealthy than, on average, individuals in sample A.

The data collection instruments

The children in the sample were administered the TORCH reading comprehension test. (Mossenson, Hill and Masters, 1987) They were also asked to complete a writing exercise specifically designed for the evaluation of First Steps. Both are now described.

The TORCH reading comprehension test

To complete the TORCH reading test students first have to read an extract of text. They then fill in blanks in a rewritten version of the original text. The practice example from the TORCH test - completed by the children before beginning the test - illustrates the kind of tasks required of them.

The text to be comprehended in this practice example is as follows:

Septimus Potter lived all alone and that was how he liked it. Of course, there was his black cat Tobias; but Tobias was a loner, and he was such a good hunter that Septimus didn't even have to feed him very often. Sometimes, when he found feathers scattered around the lily pond, Septimus cursed Tobias; but otherwise they lived in silence.

The student then tries to correctly fill in the gaps in the following passage:

The passage is about a man called _____. He lives with his _____, who sometimes upsets his master because _____

Scoring the TORCH test

The test is scored by first counting the number of correct responses. The TORCH manual provides a set of correct responses for each blank space. When a student responds with a phrase that is not listed in the manual it is judged on its correctness according to how similar in meaning it is to the list of correct responses. The manual also lists what are regarded as incorrect responses. As TORCH is a test of reading comprehension and not of spelling or of grammar, responses that have the correct meaning but which have spelling or syntax errors, are scored as correct. If the syntax or spelling is so bad that the meaning of the response cannot be established then the response is treated as incorrect. Once the raw scores are added up, they are

transformed onto a scale that provides an estimate of the reading comprehension ability of the student.⁸

This scale provides data at an interval level of measurement. These data can be legitimately summarised using the mean as the measure of central tendency and the standard deviation as the measure of spread.

Selection of the texts used in the TORCH test

In choosing a text to use with the TORCH test it was important to select one that was difficult enough to discriminate between students. If the text was too easy then many students would score all answers correct and so it would not be possible to discriminate between them. However, consideration had also to be given to the impact of a too difficult test on the self esteem and educational development of the children. To overcome this problem, before the TORCH test was administered, each classroom teacher participating in the evaluation was sent a selection of two texts taken from TORCH. Teachers were asked to indicate which of the two texts would be the more appropriate for their class. They were asked not to select a text that was too easy or too difficult. The two texts which teachers were asked to select from were 'Feeding Puff' and 'The Cats'. (See Appendix B. to examine these passages.) These two texts were chosen because the TORCH manual suggested that the ranges of difficulty of these two passages were appropriate for the ability of Year 5 students. Of these two texts 'The Cats' was the more preferred because its range of difficulty is consonant with the full range of reading comprehension abilities at the Year 5 level. The easier text, 'Feeding Puff' does not, on the evidence from the TORCH manual, capture the reading comprehension abilities of the most capable Year 5 students. (See Appendix C. which provides a graphical display of the relationship between the texts used by TORCH and the ability levels of students.) All students were administered the TORCH test using the text that the teacher chose as the more appropriate.

Two teachers felt both texts were too difficult for their class. For these classes, the text 'Lizards love eggs' was chosen. (See Appendix B. to examine this passage.)

One of the advantages of using the TORCH test is that scores from different texts can be placed along the same scale used to estimate reading comprehension ability. This was done with the data collected for this evaluation.

⁸The units of this scale are 'logits' or the natural log of the odds of a student correctly answering an item on the test. The location of an individual on this scale is the point at which it is estimated that the individual has a 50% probability of correctly answering an item from the set of items ranked in order of their difficulty.

The number of valid test papers returned showing the distribution of the different texts across the strata of the sample is displayed in Exhibit 2. This shows that about 75% of students completed the preferred text 'The Cats', about 22% of students completed 'Feeding Puff' and about 2% of students completed 'Lizards love eggs'.

Exhibit 2: Number (weighted) of valid TORCH tests returned by strata in sample.

	The Cats			Feeding Puff			Lizards Lay Eggs		
	NonFS	NewFS	OldFS	NonFS	NewFS	OldFS	NonFS	NewFS	OldFS
PSP	39	6	11	33	9	39	0	0	0
NonPSP	434	19	0	33	9	21	0	0	13
Total	472	25	11	66	18	60	0	0	13
Row %	71	4	2	10	3	9	0	0	2

(Note that rounding errors occur in this table.)

Teachers' responses to the use of the TORCH test

After TORCH had been administered to the children, teachers were asked to comment upon any inadequacies that they felt the TORCH reading test had as an instrument for measuring the reading comprehension skills of the students in their class. Of the 31 classroom teachers who returned completed TORCH test sheets, seven provided comments about the adequacy of the test. (Most provided more than one comment.) The most frequent comment was about the inappropriate difficulty of the text used. This was felt to lead to some students performing below their ability. Six teachers claimed that this was a problem. Three teachers commented that the print sizes of the text or the format of the test were unfamiliar to students and that this would affect their ability to demonstrate how well they could read. Three teachers reported that some children, when confronted early on in the test with an item perceived to be difficult, gave up hope of getting any correct and so did not complete the test. The reading comprehension of these students may have been better than their test scores indicated. Two teachers felt that the last week of the school term was a bad time to administer the test. They implied that because of this the scores of their students on the test would be underestimated. Two teachers complained that the children were not permitted to read their answers out aloud. This was the normal process of review in their class. One teacher argued that the TORCH test failed to address three important levels of questioning. These levels were described as: (1) literal answers, (2) inferential answers and (3) appreciative or 'bring your own answers'. Because of this, the teacher felt that her children were not able to demonstrate fully their reading comprehension skills.

The distribution across each of the strata of teachers making these observations about the limitations of TORCH was somewhat uneven. Exhibit 3 shows this.

Exhibit 3: Number of teachers in each stratum reporting that the TORCH test or the conditions under which the TORCH test was administered led to the reading comprehension ability of students being underestimated.

	Disadvantaged (PSP)			Non Disadvantaged		
	Non FS	New FS	Old FS	Non FS	New FS	Old FS
n of teachers reporting under-estimation	1	3	2	1	0	0
n of all teachers	5	7	7	3	4	5

Nearly one half (3 of 7) of the teachers from new First Steps disadvantaged schools reported that the students in their classes had their reading comprehension ability underestimated by the TORCH test or by circumstances prevailing at the time of the test. More generally, 6 of the 7 of the teachers reporting underestimation of their students' reading ability came from disadvantaged schools.

If the claims that these teachers make are correct then it is probably the case that the reading ability of at least some of the children in the sample is under-estimated. In particular, the reading comprehension ability of students from new First Steps disadvantaged schools and more generally from disadvantaged schools and hence for students of low socio-economic status background, may be underestimated.

Most tests were administered in the last week of the third term⁹ so it can be reasonably assumed that the distractions experienced by the students at this time were similar across all schools in the sample.

The writing exercise

The writing 'exercise' proved to be complicated in all phases of both its administration and scoring. These differences arose partly because there was no satisfactory 'off the shelf' test of writing skills for Year 5 students available. A 'test' had to be devised for the evaluation. It was decided to ask students to submit a report about an animal of their choice. The report was to be drafted under test conditions. Then, so the normal procedures used by the children to review

⁹A small number of teachers administered the test in the early weeks of the fourth term because of unsuitable conditions in the week when the test was asked to be administered. These conditions included a 'flu epidemic in the school, school camps and an end of term Mardi Gras.

and edit their writing could be examined, they were asked to write a final draft using these normal procedures. It was hoped that by doing this, evidence would be collected both about the processes of writing used by the children and the quality of their writing products. While this approach provided some fascinating qualitative material (not reported here) it did make scoring of the writing difficult because it was not always clear how much of the writing was done by the individual student and how much was due to help from peers and teachers. (See Appendix D. for the forms used in the administration of the writing exercise.)

Scoring of the writing exercise

Scoring the writing exercises involved consideration of both the draft and redraft. Each was read and judged by 2 readers. It was graded by using *The ACER Criteria for the Assessment of Writing*. (See Appendix E.) The grading was based on norm-referenced expectations. Each reader aimed to give about 30% of the students a grade of 3 out of 5, 25% of students 4 out of 5, 25% of students 2 out of 5, and 10% of students 5 out of 5 and 10% of students 1 out of 5. (These proportions were regarded as an ideal and not as a requirement.)

The readers assessed the students' writing ability by trying to separate the students' performance from the contributions of peers and the teacher. Teachers had been asked to indicate, by using different coloured inks, where they or the students' peers had contributed to the writing. Where this was done it was relatively easy to identify the unique contribution of the students. But it was not always done. As well, even when different coloured inks were used, the readers grading the texts were probably influenced by the additions of peers or teachers.

Scripts were assigned to readers in schools groups. Groups of scripts were not randomly assigned to readers although they were taken off a pile that had no planned order. Readers selected one school group. They read and judged the scripts then placed them on another pile ready for another reader. They then took another group of scripts from another school and continued until all scripts had been processed. At the end of the second grading, the scores of students allocated by both readers were compared.

Discrepancies in the scores between two readers were reconciled. Adjacent scores (5 and 4, or 2 and 1, etc.) were judged to be consonant, but separation of more than one grade (5 and 3, or 3 and 1, etc.) were judged to be discrepant. Adjustments were made by the two readers justifying the grade they had awarded and, based on arguments about these justifications, one or both of the scores were changed. There were 78 pairs of scores (about 11% of all valid cases) which had discrepant scores. It is these adjusted scores that are used in this report. A student's final score was obtained by summing together the scores from each of the two readers.

Because of these unexpected complexities in scoring the writing exercises, only the scores attached to the product of the student's work are examined in this report. The processes revealed in the work of the students are not, therefore, examined in detail. Some information is available, however, because teachers were asked to describe the process of review that their students undertook when writing. The teachers' responses showed that a wide number of approaches were used. However, it was not always clear whether the sometimes limited number of strategies reported as being used by children in some classes was their normal classroom practice or was a practice adopted to approximate to 'test conditions' for both the draft and the rewrite of the writing exercise. All teachers reported that children edited for spelling and punctuation and reread (aloud or silently) for meaning. Many also reported peer review, teacher assistance, use of dictionaries (personal, class and others), use of 'Have-a-go' pads and one teacher reported that parents helped in the classroom.

Teachers' responses to the writing exercise

Some teachers were critical of the writing exercise. They were not asked to comment upon the worth of the exercise so it is not possible to gauge how widely spread the unease was amongst them. It is interesting, however, that the most strident objections came from old First Steps schools. Among these criticisms was the concern that test conditions were inappropriate for the drafting stage of a report. It was argued that it is at this stage in preparing a report that research and planning are undertaken and so writing is not the main focus. Consequently, the children were not given the best opportunity to write. As well, it was felt by some teachers that the exercise was neither purposeful nor relevant to their children.

DATA ANALYSIS

Limitations of the data

The effect of changed circumstances on how the data are analysed

The research was designed to examine change in the reading ability of students over time and not to examine differences in the reading ability of students at one point in time. It is very unlikely, however, because of changed circumstances¹⁰ that further TORCH reading ability and writing ability data will be collected for the evaluation of First Steps. If this is so, then there is an obligation to use those data already collected as best they can be. To do this, the available data will have to be treated as if they had been collected to measure differences between groups at one point in time. This creates a number of limitations. Much of the following over the next few pages is concerned with explaining these limitations.

The discussion of the limitations begins by examining the underlying logic of the data analysis used here.

The underlying logic of the data analysis

The analysis of TORCH score data is based upon an examination of the mean or average scores achieved by various groups of students. Had these data been collected using a design which permitted a pre-test post-test comparison, then the differences between the means would have been examined with a statistical procedure called ANOVA. ANOVA will not be used for the analysis of the data here,¹¹ but the logic of the data analysis that is used will parallel the logic of the ANOVA procedure. (For readers unfamiliar with ANOVA, see Appendix H.)

The objective of the data analysis is to establish the main effects of First Steps and of socio-economic status background and the interaction effects of these two variables or factors.

In the case of the TORCH data, the interaction effect is understood as the extent to which the average TORCH scores of students in each of the three levels of involvement with First Steps are affected by their socio-economic status background.

A telling method of examining the main effects and the interaction effect is to graph the means. (The graphs are especially important in this report because they will form the basis of the data analysis.) Some explanation will help to interpret these graphs.

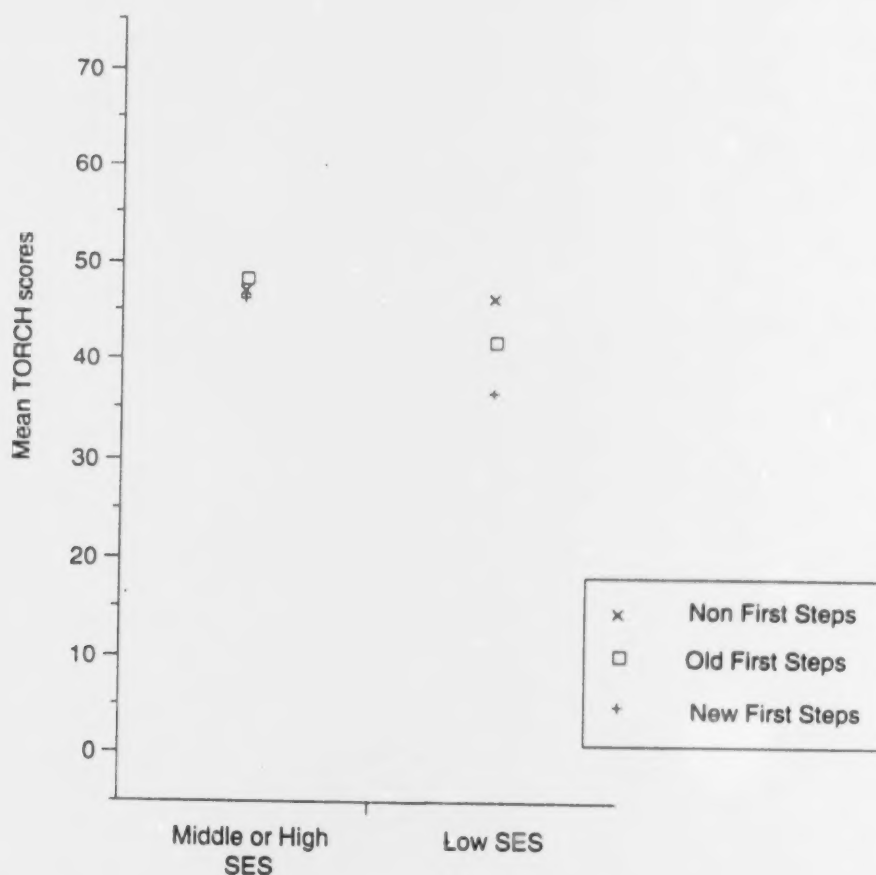
¹⁰These circumstances have to do with reduced amounts of money.

¹¹See Appendix G. for why tests of statistical significance are not used.

Graphing the main effects and the interaction effect

The graph used to depict the main effects and the interaction effect in the TORCH data consists of six points located in a two dimensional space. The six points represent the six mean TORCH scores for each of the groups of students formed by the strata in the sample. These points are located by reference to the scale used to measure the mean TORCH score on the vertical axis and to the scale used to measure the socio-economic status background of the students on the horizontal axis. Exhibit 4 is a graph of these data.

Exhibit 4: A rudimentary graph depicting the mean TORCH scores for students from the six strata in the sample.

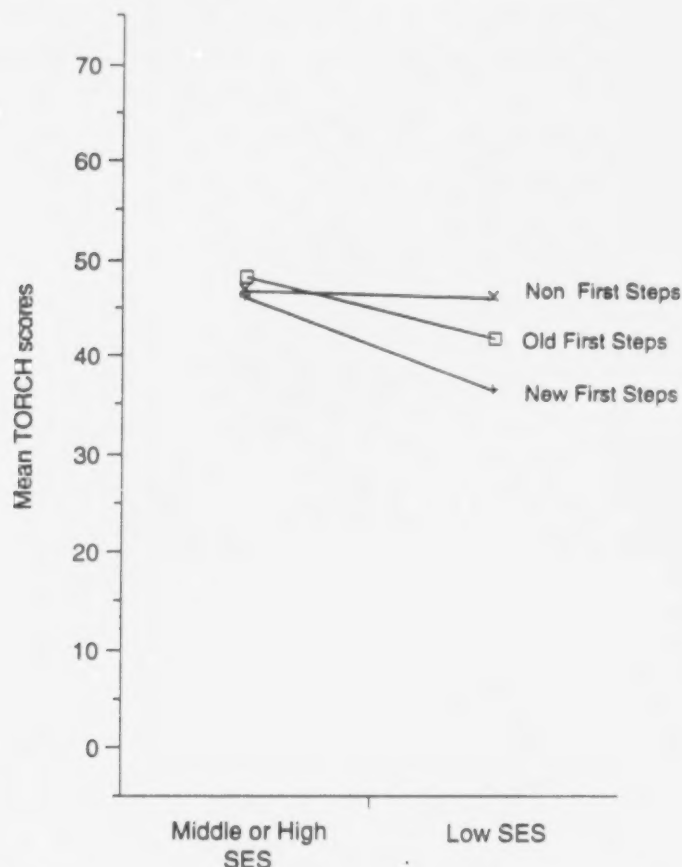


Each of the three points in each column of the graph represents either the mean score for students from new First Steps, non First Steps or old First Steps schools. They are grouped into columns according to whether the students come from a low or a middle or high socio-economic status background. To assist in the interpretation of the data displayed in this graph

the points can also be joined by a line so that the differences between means are clearer.

Exhibit 5 shows the data points joined by lines.

Exhibit 5: A complete graph depicting mean TORCH scores for students from the six strata in the sample.



The slope of the lines can be used to assist in the interpretation of the data. For example, in Exhibit 5 by using these lines as a guide, it is clear that with this set of data, the scores of students from non First Steps schools (joined by a near horizontal line) have about the same mean score irrespective of whether they have a low socio-economic status background or a middle or high socio-economic status background.

The effect of the changed role of the data on the interpretation of the TORCH reading comprehension scores

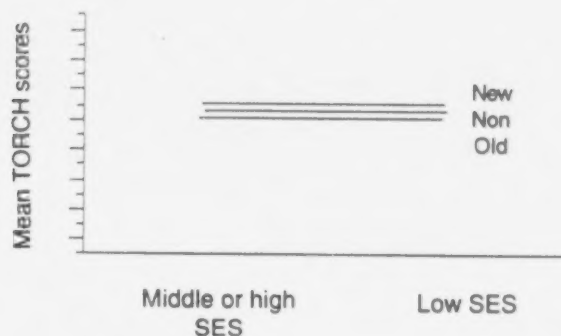
This section of the report describes the likely effect on the average TORCH scores that the revamped design of the research has had. (The effect of this revamped design is called here the 'research design effect'.)

It will be recalled that an ANOVA tests the hypothesis that the samples of scores come from the same population of students and so it is expected that the differences between the mean scores will be about zero. However, it is not the case, with the data reported here, that the expected differences between the means will be near to zero if there are no effects in the data. This is because of the research design effect. This section explains why this is so and the implications that this has for the interpretation of the data.

The effect of the research design is explored first by considering how the data would appear without this effect. It begins by showing how the mean scores with no main effects and no interaction effect would appear and then how each of the main effects would appear without the 'research design effect'. This is done to provide a comparison with the set of observed effects and so make the implications of the 'research design effect' easier to understand.

If there were no main effects, no interaction effects and no 'research design effect' in these data then there would be little difference among the six sets of mean scores. Exhibit 6 shows how data with no effects would probably appear.¹²

Exhibit 6: Expected location of mean TORCH scores for each stratum in the sample assuming that there are no effects in the data. (Order of mean scores for each level of involvement in First Steps is arbitrary)



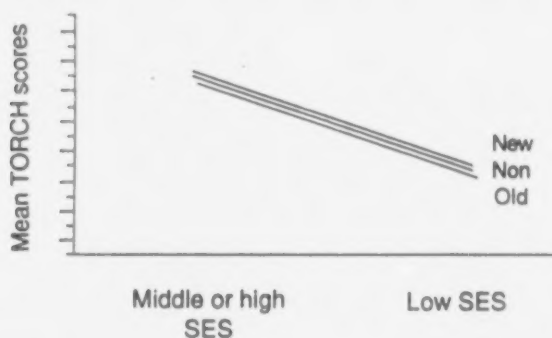
In this exhibit the mean score for each of the three levels of involvement with First Steps is at the about the same level so the lines linking each of the pairs of points between the two levels of socio-economic status background are (1) horizontal, (2) very close to each other and (3) about parallel to each other. Were there to be no random variation, all means would be equal and so each of the three lines seen in Exhibit 6 would overlay each other. However, it is

¹²Note that the Y axis does not have a scale marked upon it. While in this case, it is possible to estimate the likely location of the data points in all subsequent graphical depictions of expected effects in the data it is not possible to provide such information. There are insufficient data to establish these values. The important features of Exhibits 6 to 10, 12 and 13 is the location and the slope of the lines relative to each other and the order of the mean scores for each of the three levels of involvement with First Steps. The implications of not knowing the value of the means is discussed at length on pages 20ff.

unlikely that there would be no random variation, so the three lines have been drawn close together and are only about parallel. Since there are no effects, the order in which the points are labelled in Exhibit 6 is arbitrary.

Exhibit 7 shows what would be expected if the socio-economic status background of students had an effect on the mean scores of students.

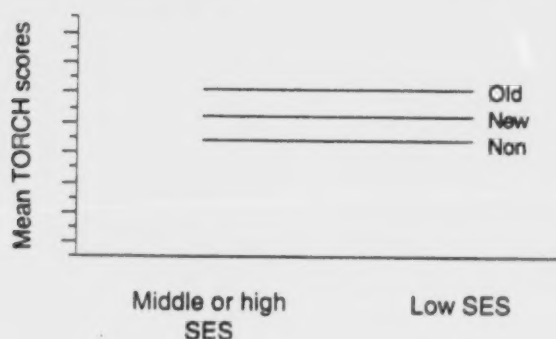
Exhibit 7: Expected location of mean TORCH scores for each stratum in the sample showing a likely main effect of the socio-economic status background of students. (The order of the mean scores for each level of involvement in First Steps is arbitrary)



It would be expected that the mean scores of students from low socio-economic status backgrounds would be lower than the mean scores of students from middle or high socio-economic status backgrounds. As the three levels of First Steps are not having an effect in Exhibit 7, there is little difference between the mean scores within each of the two levels of socio-economic status. (For example, all children from non First Steps, new First Steps and old First Steps schools within the low socio-economic status background group have about the same mean score.) Consequently, in Exhibit 7 the lines (1) slope downwards (2) are close together and (3) are about parallel. In this exhibit the order of the labels for the level of involvement with First Steps is arbitrary.

Exhibit 8 shows what could be expected if First Steps had had an effect on the mean TORCH scores of students. It shows that, all else being equal, if First Steps was having an effect then those most exposed to the program (old First Steps students) would score higher than those with less exposure to the program (new First Steps students) who would in turn score higher than those never exposed (non First Steps students) to the program. Thus it could be expected that the lines in Exhibit 8 would be (1) about horizontal and so (2) about parallel to each other and (3) ordered such that old First Steps school had the highest means score and non First Steps schools had the lowest mean TORCH score.

Exhibit 8: Expected location of mean TORCH scores for each stratum in the sample showing a likely main effect of level of involvement with First Steps. (The order of the mean scores for each level of involvement in First Steps is *not* arbitrary)



These then are the three simplest sets of effects that could be found in these data. It is now time to consider the effect of the research design and how this modifies the expected effects that have just been described.

So far the examination of the expected effects has proceeded on an assumption that schools were randomly assigned to their level of involvement with First Steps within each of the two socio-economic status background groups. As a consequence, it has been assumed that those students within the low socio-economic background group, who have had most exposure to First Steps (ie old First Steps schools) would have had, before First Steps was introduced, levels of literacy that were the same, on average, as those students of low socio-economic status background from new and non First Steps schools. In fact this assumption is probably wrong. The level of involvement of a school with First Steps was determined by the level of need of the school for a literacy program.¹³ Thus old First Steps schools are those schools that were seen as having the most need, and non First Steps schools as having the least need. New First Steps schools were seen as having a moderate need for the program. Given this, before the First Steps program was instituted it would be expected that the mean TORCH scores for each of these groups of schools would be in the following order:

non First Steps schools
new First Steps schools
old First Steps schools

with non First Steps school students having the highest mean score and old First Steps the lowest. It is this ordering of the data that constitutes the research design effect. Exhibit 9 shows the expected effect of the research design. Note that Exhibit 9 differs from Exhibit 8

¹³This information was provided by First Steps personnel.

(which shows an hypothesised main effect of First Steps) in the order of the three levels of involvement with First Steps.

Exhibit 9 Expected location of mean TORCH scores for each stratum in the sample showing the effect of the research design (The order of the mean scores for each level of involvement in First Steps is *not* arbitrary)

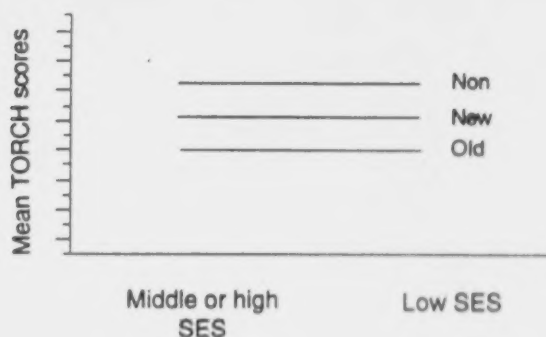
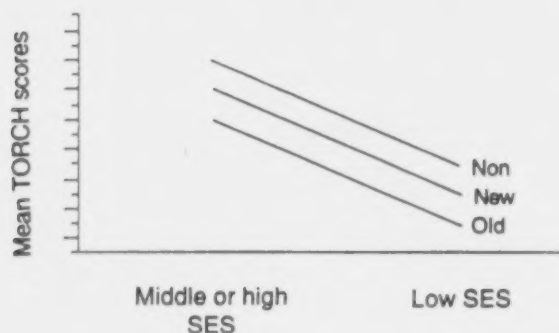


Exhibit 9 shows only the research design effect. It does not show any main effects nor an interaction effect. If the data formed this pattern then it could be concluded that there was probably insufficient evidence to claim that First Steps was affecting the reading comprehension ability of students. (It is inconclusive evidence, however, because it is not known how much the First Steps program could have changed the location of the mean scores without changing the order of the three groups. This is discussed in more detail below.¹⁴)

Exhibit 10 shows what would be expected if the socio-economic status background of students was a main effect while taking into account the research design effect.

Exhibit 10: Expected location of mean TORCH scores for each stratum in the sample showing the expected main effect of socio-economic background taking into account the effect of the research design. (The order of the mean scores for each level of involvement in First Steps is *not* arbitrary)

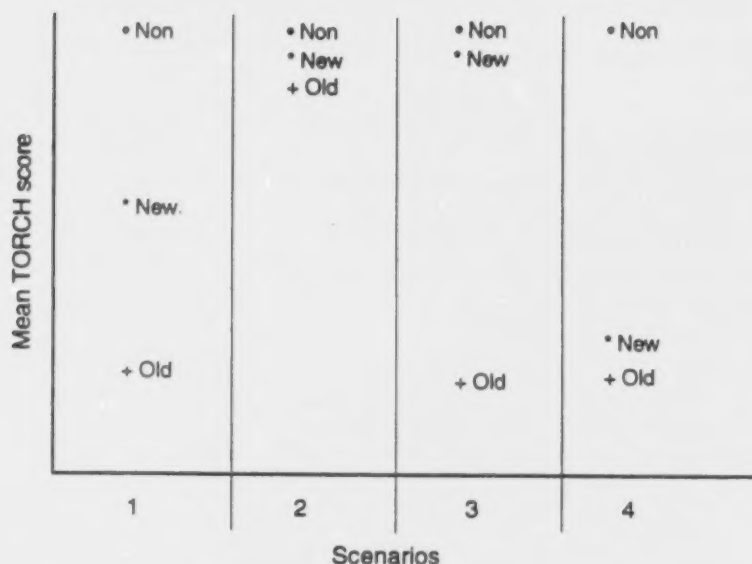


¹⁴See pages 20ff

Exhibit 10 shows how the order of the mean scores for each of the three levels of involvement with First Steps remains unchanged but now the low socio-economic group is seen to be scoring lower on average than the middle or high socio-economic status group. If this pattern was observed in the data then it could be interpreted that probably only socio-economic status background and not First Steps was affecting average TORCH reading scores.

So far it has been possible to mark out, even if only roughly, the expected effect of the research design and of socio-economic status background on the mean TORCH scores of the various groups of children. It has not been possible to describe the values of these likely mean scores nor, therefore, the likely differences between the means. There are insufficient data to do this. When turning to consider the main effect of the level of involvement with First Steps, taking into account the effect of the research design, this lack of data causes a problem. Consider Exhibit 11 which depicts four different, but possible, scenarios. Each scenario shows some of the many possible values for mean TORCH scores for each of the three groups of students before the introduction of First Steps. (Socio-economic status background is ignored in Exhibit 11.) The exhibit shows that there could have been large or small differences between the means. There could also have been equal or unequal differences between the means.

Exhibit 11: Four possible scenarios depicting possible differences between mean TORCH scores before the introduction of First Steps.

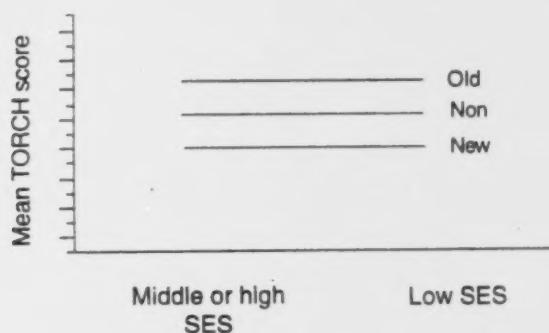


In Scenario 1, in Exhibit 11, the mean scores for each of the three levels of involvement are far apart. This means that if this scenario depicts the average reading ability of students before First Steps was implemented then there was, on average, a large difference between the reading

abilities of the three groups of students. In Scenario 2 the mean scores are close together. In Scenarios 3 and 4, the gaps between the new First Steps and the old First Steps groups varies as do the gaps between the new First Steps and non First Steps groups. It is not known which, if any, of these scenarios depict the relationship between the average TORCH scores of the groups before First Steps was introduced. As a consequence, it is very difficult to validly establish if First Steps has had any effect on the reading ability of the students. For example, consider the possibility that the mean scores of the three groups after the introduction of First Steps were found to be like that displayed in Scenario 2 of Exhibit 11. If Scenario 1 in Exhibit 11 depicts the differences between the means before First Steps, then an outcome like that seen in Scenario 2 would be evidence that First Steps was having a powerful effect. On the other hand if Scenario 2 also represented the differences between the average scores before the introduction of First Steps then to have the same outcome after First Steps would count as evidence that First Steps was having little effect on the reading ability of students.

While it is not possible to gauge the importance of the differences between the mean scores of each group of students there is one way which allows a guess at the effects of First Steps. If the order of the mean scores changes from that depicted in Exhibit 11, then this will count as some evidence for an effect in the data. For example, if old First Steps students are found to have a mean score that is higher than new First Steps students then this would suggest that First Steps could be having an effect. It is not possible with these data to establish how big or important this effect is, but nevertheless this will count as some evidence that First Steps is having an effect.¹⁵ Exhibit 12 shows one possible set of mean scores which would count as evidence, with these data, that First Steps was having an effect.

Exhibit 12: Effect of First Steps taking into account the research design effect. (The order of the mean scores for each level of involvement in First Steps is *not* arbitrary)



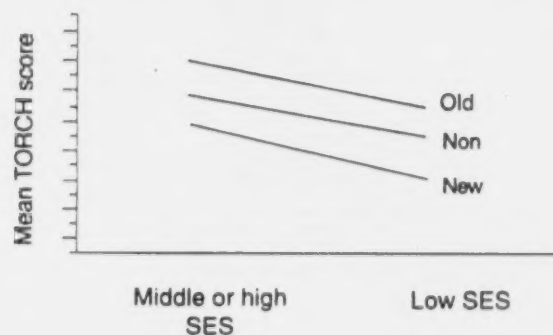
¹⁵It is possible that the change in the order has been caused by random variation in the groups of students under study. A test of statistical significance is usually used in this situation to establish the likelihood of this occurring. No such tests are used with these data. For an explanation see Appendix H.

Note that it is the order of the level of involvement with First Steps which is important here. The distance between the means is unknowable with the current data.

One interpretation of Exhibit 12 is that those students most exposed to First Steps have on average caught up with and then exceeded the average reading ability of students not exposed to First Steps. Those new to the program have yet to do so. This, in Exhibit 12, has occurred irrespective of the students' socio-economic status background. (In other words, there is no main effect for socio-economic status background.) Note that if the means for students at non First Steps and old First Steps schools were about the same, then this could also tentatively count as evidence that First Steps was having an important impact on the reading comprehension ability of students.

Exhibit 13 shows the effect of both socio-economic background and First Steps taking into account the research design effect. Note that the lines are about parallel which indicates that there is no interaction depicted in this exhibit.

Exhibit 13: Main effects of socio-economic status background and level of involvement with First Steps taking into account the research design effect. (The order of the mean scores for each level of involvement in First Steps is *not* arbitrary)



It has been demonstrated that it is possible to make informed guesses of what the effect of First Steps using the present set of data. It has also been shown what the likely effect of socio-economic status background will be. No attempt has been made to examine the likely interaction effects taking into account the research design effect. There are too many imponderables for this to be useful. Nevertheless, it has been possible to sketch out what an outcome will look like which shows that First Steps may be having an effect. It is now time to explain why only this is possible and why it is not advisable to use tests of statistical significance to guide the data analysis.

A preliminary note: statistical significance versus practical or theoretical importance

The data used here are taken from a sample of students (and not the population of all WA government Year 5 school students). Whenever a sample is taken, there is a possibility that the data are not truly representative of the population. This means that any questions asked about the differences found between groups of persons represented in the sample have to be worded in terms of probabilities or likelihoods. Thus the form of the question which is usually asked of such data taken from a sample is as follows:

"Is the difference in the average scores between the groups (within the sample) so unlikely to occur by chance that this difference is likely to be also observed in the group of all persons?"

If the difference between the two groups in the sample is such that it is very unlikely to have occurred by chance then this difference is said to be 'statistically significant'. A statistically significant difference, however, does not necessarily mean that there is a big difference. A statistically significant difference is a difference that is likely to be 'real' or, in other words, to be found in the group of all persons from which the sample was drawn. Having established that a difference is likely to be real, the next task is to establish if the difference is, in practical or theoretical terms, consequential. This step - of establishing its importance - necessarily involves making a value judgement. For example, if it was shown that a group of students scored on average 50% of items correct on a mathematics test and another similar group, which had been exposed to a new mathematics program, scored on average 95% of items on the same test correct then it is likely that most people would agree that this was an important difference. But if the difference was between 50% and 60% would this still be seen as important? What about between 50% and 55% or between 50% and 51%?¹⁶ At some point the judgement would be made that some of these differences are not important even if they can all be shown to be statistically significant. Statistics cannot assist in making this judgement.

Thus, while statistical significance is important for establishing the likely reality of observed differences, it does not indicate the theoretical or practical importance of those differences. It is important to understand this because in this report tests of statistical significance are not used. (The reason for not using a test of statistical significance is described in Appendix 11.)

¹⁶It may also not just be a matter of the size of the difference that is important. For example if a statistically significant difference was found between two groups which represented say a 5% improvement and if the expenditure on the 'treatment' program was \$100 then this program is more likely to be seen positively than if the program had cost \$1 million.

It will be recalled that when samples are drawn the questions are asked of these data ought to be couched in terms of probabilities. It is intended to still couch the analysis in terms of probabilities. The difference will be that, because tests of statistical significance are not being used, the probabilities will not be quantified.

Organisation of the data analysis

The data analysis will be presented in the following order:

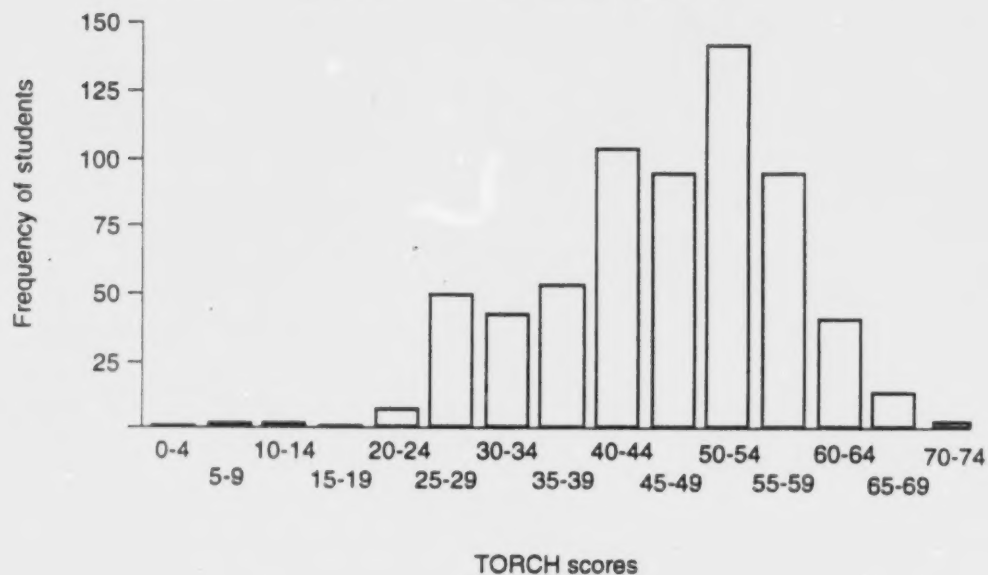
- Examination of the distribution of TORCH reading comprehension scores
- Examination of the distribution of writing ability scores
- Analysis of TORCH scores
- Analysis of writing exercise scores

Distribution of TORCH reading comprehension scores

The median of the scores is 44 and 50% of them fall between 35 and 52. The interquartile range is therefore 17. (The mean score is 46.0 with a standard deviation of 10.9.)

Exhibit 14 shows the distribution of scores achieved by students on the TORCH test. It shows that the distribution is somewhat asymmetrical with a long tail trailing into the low scoring end of the scale.

Exhibit 14: Distribution of TORCH reading comprehension scores in the total sample (weighted) of Year 5 WA government primary and District High School students.



There are a variety of reasons that might explain the observed distribution of scores. The comments made by some teachers about the limitations of TORCH may be pertinent here. If all or most of the students in the classes reported by these teachers had their reading ability under estimated, then some 15 to 20% of the students in the sample would have low scores.¹⁷ This would significantly effect the distribution of the scores, skewing them in the observed direction. Children prematurely giving up because of the perceived difficulty of the test could account for the very low scores at the end of the tail in the distribution.

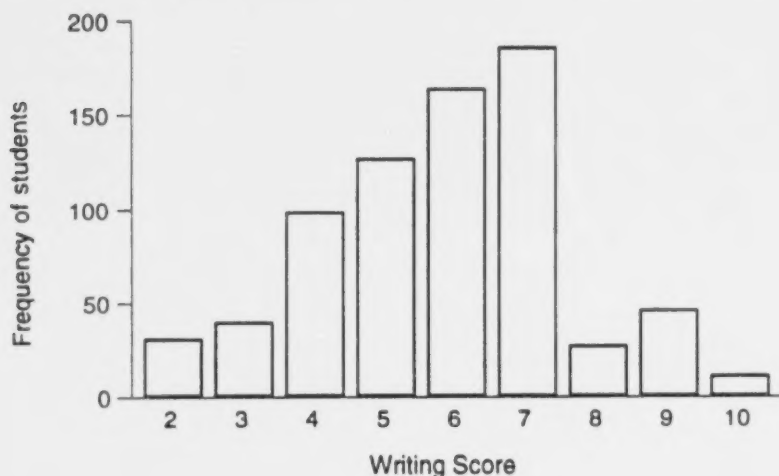
It is worth noting that of those students with a score below 20, only about 10% were identified by their teachers as remedial students and another 10% were identified as English as Second Language (ESL) students. About 70% of these students came from disadvantaged schools.

Distribution of writing exercise scores

The median of the writing scores is 5. The interquartile range is 3. (The mean score is 5.8 with a standard deviation of 1.76.)

Exhibit 15 shows the distribution of scores achieved by students on the writing exercise. The distribution appears asymmetrical given that the readers who scored the writing exercise were aiming to allocate scores ~~where~~^{which} were about symmetrical in their distribution.

Exhibit 15: Distribution of writing exercise scores in the total sample (weighted) of Year 5 WA government primary and District High School students.



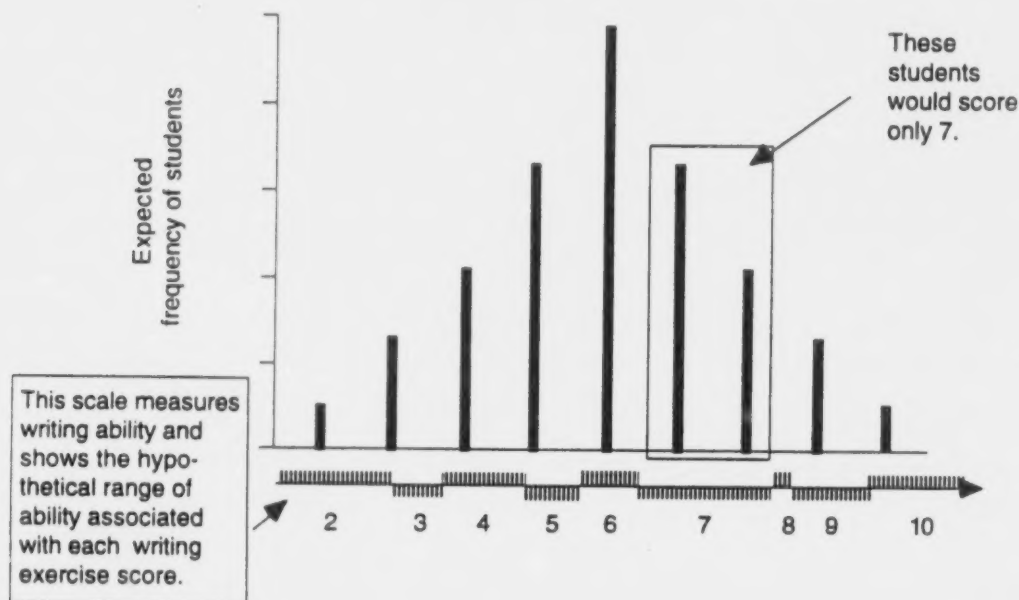
In Exhibit 15, it can be seen that the score of 7 seems over represented and the score of 8 under represented. An examination of the frequency of the scores 3 and 4 (the scores awarded by

¹⁷It is not possible to establish this figure more precisely because no record was kept of which specific classes within a school participated in the TORCH tests.

each reader required to obtain an aggregate score of 7 or 8) showed that there was little difference between the readers allocating these marks. This suggests that the observed distribution is unlikely to arise from biased scoring by any one of the readers who judged the exercises.

One explanation for the observed distribution of writing ability scores is that they are not evenly distributed along the scale measuring writing ability. Exhibit 16 shows a hypothetical example of how the scores may be distributed along such a scale.

Exhibit 16: Expected distribution of writing scores and the hypothetical location of writing exercise scores along a scale measuring writing ability



If the actual writing ability of the children was distributed symmetrically, and the difficulty of the categories distributed as shown along the writing ability scale in Exhibit 16, then there would be many children in the wide range covered by a score of 7. Because of this wide range on the ability scale, the numbers of students with a score of 7 would seem to be disproportionately large. It should also be noted that there are no data currently available to test what is the relationship between the different categories of score for the writing exercise and levels of writing ability.

The examination of the distribution of the writing scores has shown that the relationship among the scores has to be interpreted cautiously. It is not clear that a score of, say, 5 is as different from a score of 6 as a score of 6 is different from a score of 7. This contrasts with the TORCH

reading comprehension scores, where the relationship between the scores and ability levels is known. In the case of the TORCH scores the difference between, say, 42 and 43 is the same as between 43 and 44.

The main implications of this examination of the distribution of the TORCH reading comprehension scores and the writing ability scores are as follows:

- As little is currently known about the location of the writing exercise scores along the scale measuring writing ability (except for the order) it can be expected that the TORCH data will provide a more precise picture of the effects of First Steps than will the writing exercise data.
- Summarising the writing score using the mean as a measure of central tendency is open to attack. It can be argued that the writing exercise scores should properly be treated as only ordering the ability of students.

Analysis of TORCH reading comprehension scores

The mean TORCH reading comprehension scores are displayed in Exhibit 17.¹⁸ There are a number of striking features in this exhibit. First there seems to be marked differences between the spread of the mean scores of the 'middle or high' socio-economic status background group compared with the spread of the mean scores of the low socio-economic status background group. Secondly, because the three lines are not parallel, there seems to be an interaction in these data. Thirdly, the order of the three levels of involvement with First Steps for the low socio-economic status background groups (non, new and old) suggests that First Steps has affected the reading comprehension ability of students. (Exhibit 18 shows the order of the mean scores as they would appear taking into account the research design effect and the effect of socio-economic status.)¹⁹ Fourthly, the order of the three levels of involvement with First Steps in the middle or high socio-economic status background group (old, non, new) is also pointing to First Steps as having affected on the reading ability of students. Each of these four features of Exhibit 17 is now discussed.

¹⁸For an explanation of Exhibit 17 see pages 14ff.

¹⁹For arguments in defence of this see pages 21ff. Exhibit 18 is the same as Exhibit 13. It is included here for ease of reference.

Exhibit 17 Mean TORCH scores for Year 5 Western Australian government school students from the six strata in the sample.

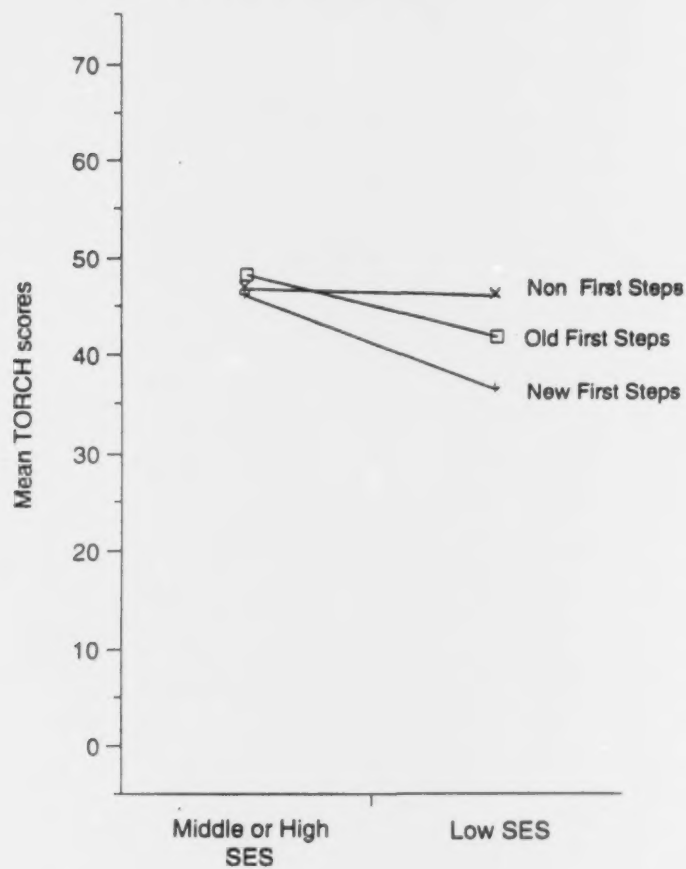
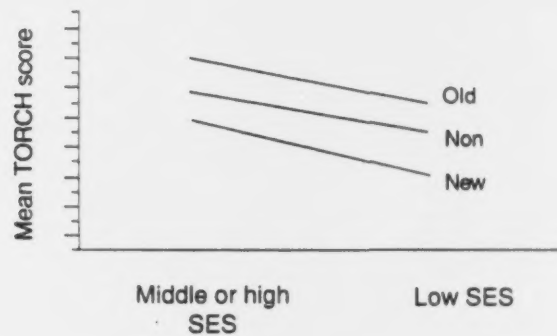


Exhibit 18: Hypothesised description of the main effects of socio-economic status background and level of involvement with First Steps taking into account the design effect.



The spread of the mean TORCH scores

While the different spread of the mean score for low and for middle or high socio-economic status background seems to be quite a striking feature of Exhibit 17, little can be made of it. This is because the mean scores prior to First Steps being introduced is unknown for each of these groups of children. (See Page 20 for a further discussion of this.)

The apparent interaction effect

It will be observed in Exhibit 17 that for students from new First Steps schools and old First Steps schools the apparent effect of socio economic background is to lower the mean TORCH scores of the low socio-economic status students. In contrast, the mean scores for students from non First Steps schools are about the same irrespective of their socio-economic status background. This is a surprising finding. It was expected that in non First Steps schools the socio-economic background of students would show up as an effect. (See Exhibit 18 for the expected effect. In this exhibit the lines slope downwards and are about parallel. In Exhibit 17 the non First Steps line is not parallel with the other lines in the graph.) Further, because First Steps was designed to assist those children who are 'at risk' it was expected that if the effect of socio-economic status background was to be reduced, it was most likely to be reduced in old First Steps schools and least likely to be reduced in non First Steps schools. (This effect is not shown in Exhibit 18) There are a number of possible explanations for this finding.

- i The small difference in the mean score for non First Steps students from a low socio-economic background compared with the mean score for non First Steps students from a middle or high socio-economic status background could have occurred because of random variation in the sample. The true means may have other values. The probability that this is so cannot be established with these data.
- ii The small difference in the means could have occurred because socio-economic status has no effect on the reading comprehension ability of students in non First Steps schools.
- iii The small differences in the means TORCH scores of low socio-economic status background students and middle or high socio-economic status background students in non First Steps schools could have occurred because students have about the same mean reading ability in these schools. Given that both types of school were seen as being of low priority for the First Steps program, (because they were not chosen to participate in the program) it is possible that the average reading levels of students are the same in non First Steps PSP and non First Steps non PSP schools.

It is not possible to establish with the available data which, if any, of these explanations account for the observed mean TORCH scores for non First Steps schools. It is therefore not possible to establish if there is an interaction effect in these data.

The order of the mean TORCH reading scores for low socio-economic status background students.

In Exhibit 17 the mean TORCH reading comprehension scores of students from a low socio economic background appear in the following order:

- non First Steps
- old First Steps
- new First Steps.

It has already been argued that the order of the mean scores for each of the three levels of involvement in First Steps can be tentatively treated as a pointer to the effect of the program.²⁰ However such a suggestion has to be also seen in the context of other, competing explanations. There are at least three possible explanations for the observed order of these mean scores. These are:

- i Students from old First Steps school had a higher reading comprehension ability than students from new First Steps schools prior to the introduction of the program. First Steps, therefore, has had no effect. This is an unlikely explanation. It is unlikely since, according to First Steps personnel, old First Steps schools were selected to participate in the program because the children in these schools had well below average literacy skills.
- ii The order of the observed means occurs because of random variation in the sample and the true means have different values. This cannot be established with the present set of data. (Note that this explanation should not be discounted as unlikely simply because there is a wide spread in the observed mean scores.) Related to this is the possibility that some student had their reading ability underestimated. It will be recalled that 6 of 20 teachers from PSP schools reported that the reading ability of their students had been underestimated by using the TORCH test.
- iii Another explanation for the observed order of the means is that the reading ability of students in old First Steps schools has improved. Further their mean reading comprehension ability has improved such that it may now exceed the mean reading comprehension ability of students in new First Steps schools. There is some anecdotal

²⁰See pages 21ff

evidence from First Steps personnel that, generally, students in old First Steps schools had literacy levels considerably lower than those observed in the new and non First Steps schools. If children from old First Steps schools have overcome this difference then this would count as evidence of the program's success.

Of these three explanations, the last two are the more plausible. It is not possible to establish which of these two is the better explanation. However, if it is assumed that, generally, First Steps has been associated with the improved reading ability of students in old First Steps schools, then with a few additional assumptions a rough estimate of the effect of the program can be made. A fair approach is to assume that, prior to First Steps, students in old First Steps had a mean reading comprehension score as far below the mean score of students from new First Steps schools as the observed mean score of students in new First Steps schools is below the observed mean score of students in non First Steps schools. On advice from First Steps personnel this is a fairly conservative assumption. If these assumptions are made and accepted, then it is possible to claim that First Steps has been associated with an increase of about 20²¹ in the mean TORCH scores of Year 5 students from a low socio economic status background. (That is, First Steps has been associated with a change of mean TORCH scores from 22.3²² to 41.9.) To help make this more meaningful, TORCH scores can be translated into percentiles. If this is done, and all the assumptions made so far continue to be accepted, then First Steps has been associated with an improvement from an average percentile of about 4²³ to an average percentile of about 37 for students with a low socio-economic status background in old First Steps schools. Using the TORCH manual as a guide to further interpret these scores, this can be seen as moving these children from a 'low' level of achievement through the 'below average' level of achievement into the 'average' range of achievement.²⁴ Such a change could rightly be regarded as an important achievement for the First Steps project.

²¹This figure was calculated as follows:

	Mean non First Steps score	46.1
less	Mean new First Steps score	36.3
		9.8

If the difference between old new First Steps and new First Steps prior to the introduction of the program is assumed to equal 9.8, then the observed mean Torch score for students at old First Steps schools has improved 9.8 to catch up to the new First Steps mean score and a further 9.8 to reach the current mean of 46.1. Hence $9.8 \times 2 = 19.6$ or about 20.

²²This figure is obtained by subtracting 9.8 from the mean score of 36.3 for new First Steps schools

²³A percentile of 4 means that 96% of all Year 5 students would be expected to score above this level.

²⁴See Appendix F. for more details.

The order of the mean TORCH reading scores for middle or high socio-economic status background students.

In Exhibit 17 it was seen that the mean TORCH reading comprehension scores of students from a middle or high socio economic background were ranked in the following way:

- old First Steps
- non First Steps
- new First Steps.

This order of the mean TORCH scores for the students with a middle or low socio-economic status background suggests that First Steps has had a positive impact on the reading ability of students. As with the data for the low socio-economic group of students this is, however, only one of a number of possible explanations for the observed order. Three explanations are now considered.

- i Students from old First Steps schools had a higher mean reading ability of students from new First Steps schools and higher than students from non First Steps schools. This is unlikely because old First Steps schools were chosen to participate in program first because of the number of 'at risk' students in them. It would be expected that they would have had the lowest mean score prior to the introduction of First Steps.
- ii Another explanation is that the means are ranked in the observed order because of random variation in the scores. The true order differs. This cannot be established with the present set of data²⁵. It should be noted, however, that the relatively narrow spread of the means for the middle or high socio-economic status group when compared with the spread of means for the low socio-economic status group does not make it any more (nor less) likely that random variation has produced the observed order.
- iii The means are ranked in the observed order because the reading ability of students in old First Steps schools has improved. Further their mean reading comprehension ability has improved such that it now exceeds the mean reading comprehension ability of students in both the non First Steps and the new First Steps schools. By making the same assumptions as were made in the preceding section²⁶ it is possible to claim that First Steps has lead to an improvement from about the 51st percentile to the 59th percentile for the mean scores of students in old First Steps schools.

²⁵A test of the hypothesis that the means are the same could be conducted, but it could not be meaningfully interpreted.

²⁶This is the assumption that the mean score of students in old First Steps schools was as far below the mean score of students at new First Steps schools as the mean of students at new First Steps is below the mean of students at non First Steps schools.

A comparison of the mean TORCH scores of the low and of the middle or high socio-economic status background students.

There are too many unknowns to devote too much space to speculating about the different effect of First Steps on the low and the middle or high socio-economic groups. However, if it is accepted that;

(1) First Steps has had an effect (as evidenced by the change in the rank ordering of the different levels of involvement with First Steps), and

(2) the difference between the mean scores of students in non First Steps schools and new First Steps schools is about the same as the difference between the mean scores of students in old First Steps schools and new First Steps schools before First Steps was introduced,

then it can be considered that First Steps has had a more powerful impact on the reading comprehension ability of low socio-economic status background than upon the middle or high socio-economic status group.

This conclusion has to be treated cautiously. It has to be treated cautiously because it is not clear what effect socio-economic status is actually having in these schools.²⁷ The conclusion also has to be treated cautiously because socio-economic status background has been defined in terms of the PSP status of the school the children attend. There could be an interaction between other programs in PSP schools and First Steps which accounts for the marked difference in the effect of First Steps for these two different groups of students.

Preliminary conclusions based on the TORCH data

The observed differences in the mean TORCH scores of students can be explained in terms of (1) First Steps having had no effect, or (2) First Steps having had an effect. If anecdotal evidence taken from First Steps personnel is accepted as valid, it is likely that First Steps has had an effect. Further, by accepting some further assumptions it is possible to argue that the effect seems stronger for low socio-economic status background children.

²⁷See pages 29ff for a further discussion of this.

Analysis of Writing exercise scores

The median scores²⁸ for the writing exercise are displayed in Exhibit 19. There are similar patterns in these data as there were for the TORCH data. As the data are subject to the same limitations as the TORCH data it will not be possible to establish if there is an interaction effect. Consequently, only the order of the three different levels of involvement with First Steps for each of the two socio-economic status groups is explored here.

The order of the median writing scores for low socio-economic status background students.

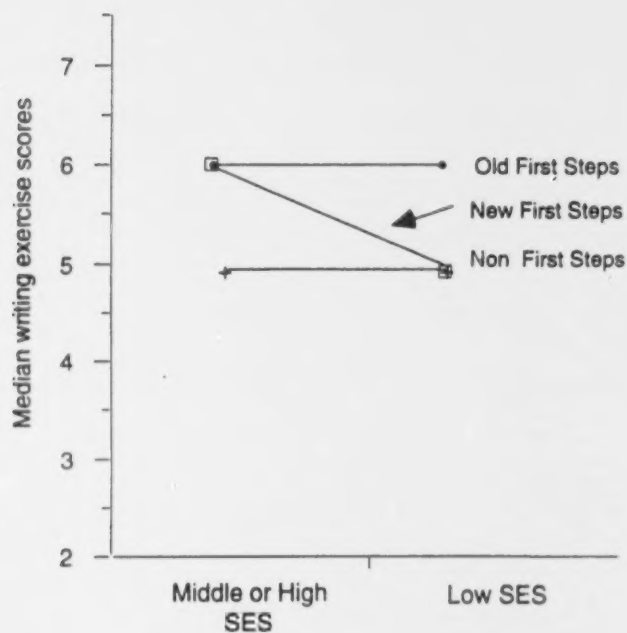
Exhibit 19 shows that for low socio-economic status students, those with most exposure to First Steps have a median score higher than students from new First Steps schools and students from non First Steps schools. This could be due to one or more of the following explanations.

- i The median scores are ordered this way because of random variation in the sample. It may be the case that the medians lay at different points for the group of all students from which the sample was drawn. These true medians may not be ordered in the same way as the observed medians.
- ii The median scores could be ordered this way because the students from old First Steps schools have always had a higher writing ability than the other students. If this is so, then it would be concluded that First Steps has had no effect on the writing ability of students.
- iii The median scores have been ordered this way because First Steps has improved the writing ability of students with most exposure to the program.

Of these three possible explanations, the second is the least likely to be true. On anecdotal evidence from First Steps staff, old First Steps schools were chosen to participate in First Steps first because of the below average levels of literacy in them. If this is the case, then these results may be treated as tentative support for the claim that First Steps has had a positive impact on the writing ability of students with a low socio-economic status background.

²⁸See pages 25ff for an explanation of why the median is used with the data and not the mean.

Exhibit 19: Median writing exercise scores for Year 5 Western Australian government school students from the six strata in the sample.



The order of the median writing scores for middle or high socio-economic status background students.

Exhibit 19 shows that for middle or high socio-economic status students, those with any exposure to First Steps have a median score higher than those students with no exposure. This could be due to the same three reasons described above for the order of the medians for the low socio-economic status group. These results can therefore also be tentatively interpreted as showing that First Steps has had an impact on the writing ability of students with a middle or high socio-economic status background.

Conclusions from writing exercise data

The results of the writing exercise suggest that First Steps is having a positive effect on the writing ability of children irrespective of their socio-economic status background. This finding is premised on the assumption that schools with most exposure to the program had the lowest level of literacy before First Steps was introduced. This finding is also premised on the assumption that the sample is not biased and that the median scores reflect the true scores of all the students from which the sample was drawn.

CONCLUSION

The following conclusions can be drawn from the data analysed here:

- i If students from old First Steps schools had lower average levels of reading comprehension ability than students in new or non First Steps schools before First Steps was introduced, then it is likely that First Steps has lead to an improvement in the reading ability of those students exposed to the program.
- ii If students from old First Steps schools had lower average levels of writing ability than students in new or non First Steps schools before First Steps was introduced, then it is likely that First Steps has lead to an improvement in the writing ability of those students exposed to the program.
- iii There was some evidence that First Steps may be making an important difference in the reading ability of students. Some informed guessing of likely mean scores suggests that the mean TORCH reading score of children in old First Steps schools may have moved from 'low' to 'average'.

BIBLIOGRAPHY

- Mossenson, L., Hill, P., and Masters, G. TORCH test of reading comprehension.
Hawthorn, Victoria, ACER, 1987
- Western Australian Ministry of Education Reading Development Continuum. Perth,
Western Australia, 1992
- Western Australian Ministry of Education Writing Development Continuum. Perth,
Western Australia, 1992

Appendix A: Weighting of the sample.

1. Stratum weights

The stratum weight was calculated first. An example will help to show how this calculation was made. The stratum weight for the respondents completing the TORCH reading comprehension test from PSP non First Steps schools was obtained as follows:

$$(10091/94021) / (94/665) = 0.75928205$$

By multiplying the achieved number of respondents (94 in the case of PSP non First Steps schools) by the stratum weight the value of 71.3725125 is obtained. This is the number of cases that would be expected if the sample contained the same proportion of cases for this stratum as the proportions in the population from which the sample was drawn.

Exhibit A1: Strata weights for the sample of students completing the TORCH reading comprehension test.

	Achieved n in selected schls	N in strata	Stratum Wt	Weighted Frequencies	Wt %
Non First Steps PSP	94	10091	0.75928205	71.3725125	10.7
Non First Steps non PSP	80	66016	5.83654715	466.923772	70.2
New First Steps PSP	154	2139	0.09823966	15.1289074	2.3
New First Steps non PSP	65	4003	0.43558109	28.3127706	4.3
Old First Steps PSP	152	7275	0.33852145	51.45526	7.7
Old First Steps non PSP	120	4497	0.26505648	31.8067772	4.8
Total	665	94021		665	100

Exhibit A2: Strata weights for the sample of students completing the writing exercise.

	Achieved n in selected schls	N in strata	Stratum Wt	Weighted Frequencies	Wt %
Non First Steps PSP	99	10091	0.79790842	78.99273567	10.7
Non First Steps non PSP	117	66016	4.41688704	516.7757841	70.2
New First Steps PSP	157	2139	0.10665079	16.74417417	2.3
New First Steps non PSP	60	4003	0.52226063	31.33563778	4.3
PSP old First Steps	180	7275	0.31638322	56.94897948	7.7
non PSP old First Steps	123	4497	0.28620072	35.20268876	4.8
Total	736	94021		736	100

BEST COPY AVAILABLE

2. School weights

Once the strata weights were calculated, the weights for each individual school was calculated. These weights were calculated by dividing the sample from each school by the mean number of students in each school for a given stratum. See below for discussion of 'final weight'.

Exhibit A3: School and final weights for the sample of students completing the TORCH reading comprehension test.

School*	n	Mean	Schl Weight	Final Weight
Non First Steps PSP: A	0			
B	44	31.3333333	0.71212121	0.54070085
C	27	31.3333333	1.16049383	0.88114213
D	23	31.3333333	1.36231884	1.03438424
Stratum total	94			
Non FS non PSP: E	0			
F	56	40	0.71428571	4.16896225
G	0			
H	24	40	1.66666667	9.72757859
Stratum total	80			
New First Steps PSP: J:	40	38.5	0.9625	0.09455567
K	51	38.5	0.75490196	0.07416131
L	23	38.5	1.67391304	0.16444465
M	40	38.5	0.9625	0.09455567
Stratum total	154			
New FS non PSP: N	0			
P	6	21.6666667	3.61111111	1.5729317
Q	51	21.6666667	0.4248366	0.18505079
R	8	21.6666667	2.70833333	1.17969677
Stratum total	65			
Old First Steps PSP: S	39	38	0.97435897	0.32984141
T	41	38	0.92682927	0.31375159
U	42	38	0.9047619	0.30628131
V	30	38	1.26666667	0.42879383
Stratum total	152			
Old FS non PSP: W	60	40	0.66666667	0.17670432
X	0			
Y	44	40	0.90909091	0.24096043
Z	16	40	2.5	0.66264119
Stratum total	120			

*School names are not used in this table because one of the conditions of the data collection was that schools would not be identified in any report of the research. The code letters A to Z are used instead. (Note that the letters 'I' and 'O' are omitted to avoid confusion in the table.)

Exhibit A4: School and final weights for the sample of students completing the writing exercise

School*	n	Mean	Schl Weight	Final Weight
Non First Steps PSP: A	0			
B	47	33	0.70212766	0.560232168
C	27	33	1.22222222	0.975218959
D	25	33	1.32	1.053236476
Stratum total	99			
Non FS non PSP: E	0			
F	51	39	0.76470588	3.377619504
G	23	39	1.69565217	7.489504118
H	43	39	0.90697674	4.00601383
Stratum total	117			
New First Steps PSP: J:	40	39.25	0.98125	0.104651089
K	52	39.25	0.75480769	0.080500837
L	21	39.25	1.88904762	0.199335407
M	44	39.25	0.89204545	0.095137353
Stratum total	157			
New FS non PSP: N	0			
P	0			
Q	53	30	0.56603774	0.295618224
R	7	30	4.28571429	2.238259842
Stratum total	60			
Old First Steps PSP: S	39	45	1.15384615	0.365057561
T	42	45	1.07142857	0.338982021
U	45	45	1	0.316383219
V	54	45	0.83333333	0.263652683
Stratum total	180			
Old FS non PSP: W	57	41	0.71929825	0.205863677
X	0			
Y	46	41	0.89130435	0.255091948
Z	20	41	2.05	0.586711479
Stratum total	123			

*School names are not used in this table because one of the conditions of the data collection was that schools would not be identified in any report of the research. The code letters A to Z are used instead. (Note that the letters 'I' and 'O' are omitted to avoid confusion in the table.)

The 'final weight is the weight applied to the individual students within each of the schools. It is the product of the stratum weight and the school weight. For example, the final weight for students who completed a TORCH test in school B was 0.75928205 (the stratum weight) by 0.71212121 (the school weight) which equals 0.54070085. The application of these weights means that each student contributes in proportion to the probability of their being drawn for the sample.

Appendix B: TORCH reading test materials

This appendix contains all the materials sent to classroom teachers for the administration of the TORCH reading comprehension test. This appendix includes the following:

- Teachers instructions for the administration of the TORCH test
- Student report form for teachers to list the names of students with special circumstances which may have affected their performance on the test
- Teacher assessment of the TORCH reading test form. This form was used by teachers to comment upon the adequacy of the test
- Text of *Lizards Love Eggs*
- TORCH answer sheet for *Lizards Love Eggs*
- Text of *Feeding Puff*
- TORCH answer sheet for *Feeding Puff*
- Text of *The Cats*
- TORCH answer sheet for *The Cats*

**Administration of the Torch Reading Test
for the First Steps evaluation, 1992.**

Notes for Year 5 Classroom Teachers.

Preparation for testing

Announcement of the testing session to students

To ensure that students are able to work at their normal levels of ability without tension or anxiety, it is unwise to place great emphasis on the importance of the tests. On the other hand, students may become upset if they receive no warning at all. Probably the best approach is to give only the kind of notice that is normally given for regular classroom tests.

When to test

The time of the day for testing will be influenced by various matters of school organization, but it is usually better to test in the morning. For valid results, tests should not be given just before or after an exciting school event.

**For the First Steps evaluation testing should occur during the week beginning
Monday September 21**

Assembling the test materials

Before testing begins, the teacher should check that the following materials are available:

- enough copies of the passage to be used in the test.
- enough answer sheets for the students for 'The Cats' (The answer sheets are attached to the passage in 'Feeding Puff')
- a supply of pencils in case of breakage;
- a watch or clock for timing the test.

Since there is likely to be variation in the time that students take to complete the test, ensure that quiet activities are available for them to go on to as they finish the test.

Remember! Test in the week beginning September 21

Practice example

- Ask the students to fill in the details above the practice example printing their first and last name, school (if required), and Year level.
- Read the text from the practice example while the students follow it on the exercise page:

Septimus Potter lived all alone and that was how he liked it. Of course, there was his black cat Tobias; but Tobias was also a loner, and he was such a good hunter that Septimus didn't even have to feed him very often. Sometimes, when he found feathers scattered around the lily pond, Septimus cursed Tobias; but otherwise they lived in silence.

- Students will probably require some support at this point. This can be provided by the teacher reading aloud the exercise and pausing to allow time for the students to write in their answers.
- Ask the students for their answers to the practice example. It is important to demonstrate to them the range of acceptable answers possible for each space. Students typically provide a wide range of answers to the practice example. The following are all acceptable answers:

The passage is about a man called - *Septimus / Septimus Potter / Potter / Mr Potter*
He lives with his - *cat / black cat / black cat Tobias / cat Tobias*
who sometimes upset his master because - *he left/scattered feathers around the lily pond / he scattered feathers around / he killed birds/ducks/geese/chickens*

- At this point it is important also to emphasize that the students should read the 'stem' both before and after the space to help them with their answers.
- Warn the students that they cannot change the printed part of the exercise in any way e.g., by altering a word, or adding an ending to a word).
- Encourage the students to make inferences based on clues in the text as illustrated by the final response given above ('he killed birds/ducks/geese/chickens').

Directions for administration

• When the students are seated comfortably in an appropriate arrangement, call them to attention and announce the name and purpose of the test. Have the students remove all materials from their desks with the exception of pencils and erasers. Distribute the test booklets and the answer sheets, saying:

"I am going to give each one of you one of these booklets. Each booklet contains a test and an answer sheet. I want to see how well you can do the test. When you get your test booklet, you are to keep it closed. Do not write anything until you are told."

• After the test booklets and, if they are separate, the answer sheets have been distributed, say:

"This is a test to find out how well you understand what you read."

On your desk you have a test booklet (hold it up). In this booklet is the passage you will read when you start the test. As well as this passage there is a page with the exercise on it."
(Show the students the exercise page.)

"You will notice that some words have been left out of the exercise. To do the test, you have to read the passage in the booklet and then complete the exercise by writing into the spaces the word or words you think are missing."

"To help you do the test, you will do a practice exercise but before you start, here are some helpful hints for you to follow:

Use a pencil and rubber so you can change your answers if you want to.

You can look back in the book as often as you like to find the answers. (Ensure that students clearly understand this point.)

Often the words you want for your answer are not in the book, so feel free to use your own words.

You can write one word answers or you can write as many words as you want in each space.

If you can't fit all of your answer in the space ask for help. (Students can dictate an answer for the teacher to write on the test exercise.)

If you spell a word wrongly it doesn't mean the answer will be marked wrong. Ask for help with spelling if you really can't write the word you want. Dictionaries are not to be used. (Teachers can supply correct spellings. It is less disruptive, however, if the children use their own spellings.)

Don't get stuck on an answer you can't fill in. Leave it and come back to it later.

When you have finished, read through the exercise to check that it makes sense.

Are there any questions?"

Starting the test

- Check again to see if the students have any questions.
- Indicate that any questions during the testing must be directed to the teacher and that otherwise there must be silence.
- Explain to students that they can work on their chosen activity as soon as they complete the test.

If different tests are to be administered during the one test session as a consequence of tailoring a test to individual students, ensure that students have the appropriate materials.

- Ask them to turn to the appropriate page in the booklet. Explain again that they are to read the passage in the booklet and then fill in the spaces in the exercise.
- Instruct students that there is no time limit for the test, and that they may keep working until they have completed the test.
- Allow students to begin the test. Note the starting time.
- If after 50 minutes students have not completed the test, allow further time only if it is apparent that they are successfully attempting items.

After the test session

- Collect all test materials from the students.
- Check to see that the answer sheets have been correctly completed. (Name, school name especially.)
- Record any unexpected variation from normal testing procedure that may have occurred.
- **Complete the teachers Test Report form.**
- Despatch **all** test materials in the envelope provided to :

Adrian Harvey-Beavis
First Steps Evaluation
ACER
Reply Paid 2
Box 210
Hawthorn Victoria 3122.

If you use this address you will not need to pay for the postage.

Permission to use Torch materials for the 1992 First Steps evaluation granted by ACER.
Copyright Minister of Education, Western Australia 1987

Name of School:

Torch Writing Test

Special circumstances

[illegible]

Please return this page with the completed test.

First Steps Evaluation, 1992:

Teacher Assessment of Torch Reading Test

Comment on any of the inadequacies of the Torch reading Test that may effect the proper understanding of the reading skills of the Year 5 students in this class.

Lizards Love Eggs

'Snake!' Tony's mother yelled.

It was the first day of the family's camping holiday.

With one hand his mother grabbed up the baby from the grass by the tent flap. With the other she seized the stick of the beach umbrella. 'Keep back!' she called as Tony ran towards the tent.

Tony laughed and ran into the tent. 'It's only a lizard, Mum. I saw it walking through the grass.'

'Are you sure?' his mother said. 'I saw the tip of its tail sliding into the tent.'

'Sure,' said Tony. 'It's a blue-tongue. See!' He pointed into the tent.

The big lizard, about forty-five centimetres long, was crawling over Tony's airbed, its blue tongue flicking.

'Quick!' his mother said. 'It'll crawl into your sleeping-bag.' She shuddered. 'I wouldn't fancy sleeping with a lizard.'

Tony tweaked his sleeping-bag and the lizard slipped down onto the groundsheet, then scabbled quickly on to the next airbed.

'Oh, no!' said Tony's mother. It was hers.

The lizard disappeared into the opening fold of her sleeping-bag. Tony pulled the bag off to shake it outside, but before he reached the door the lizard fell out. It reared its head and hissed fiercely. The baby yelled in fear at the wide gaping mouth and the long blue tongue.

'Poor fella,' said Tony as the lizard scrambled for safety into the shelter of the grocery cartons. Tony dived after it.

'If it eats the apricots, it had better look out,' Tony's mother said.

'It won't just now,' Tony said. 'We've made it nervous.'

'It nervous! How do you think I feel?' said his mother. She already felt silly about her panic in mistaking the lizard for a snake. Now she felt annoyed at seeing the tidy arrangement of her tent turned topsy-turvy as Tony hunted through it.

'Mind the eggs,' she called. But it was too late.

As Tony lunged to catch the lizard he stepped right in the middle of the egg-box.

Tony held up the lizard triumphantly. 'Look. See its goldy-pink belly.'

'Look,' replied his mother. 'See your goldy-yellow feet.'

Tony looked at the eggs oozing from the squashed box. 'Too bad it's not a goanna,' he said. 'Goannas love eggs. It would have cleaned that up in no time.'

'Too bad indeed,' his mother agreed. 'Now you'll have to do it.'

Tony put the lizard in the washing-up basin. He mopped up the squelchy mess of yolks and whites with a cloth and squeezed it into the basin. The lizard's long blue tongue flicked in and out as the thick yellow drops fell. By the time the broken eggs were cleaned up, the lizard's stomach had swollen wide and flat between its legs. Blue-tongue lizards love eggs too!



Lizards Love Eggs

Helpful Hints

Here are some helpful hints for you to follow:

- Use a pencil and rubber so that you can change your answers if you want to.
- You can look back in the book as often as you like to find the answers.
- Often the words you want for your answer are not in the book, so feel free to use your own words.
- You can write one word for your answer or you can write as many words as you want in each space.
- If you can't fit all of your answer in the space ask for help.
- If you spell a word wrongly it doesn't mean the answer will be marked wrong. Ask for help with spelling if you really can't write the word you want.
- Don't get stuck on an answer you can't fill in. Leave it and come back to it later.
- When you have finished read through the exercise to check that it makes sense.

Practice Example

Septimus Potter lived all alone and that was how he liked it. Of course, there was his black cat Tobias; but Tobias was also a loner, and he was such a good hunter that Septimus didn't even have to feed him very often. Sometimes, when he found feathers scattered around the lily pond, Septimus cursed Tobias; but otherwise they lived in silence.

The passage is about a man called _____
He lived with his _____ who
sometimes upset his master because _____

NAME.....

YEAR.....

Lizards Love Eggs

It was the first day of the family's camping holiday. Something crawled into the tent and thought it was a snake.

☐ 1

Mother grabbed up the baby because she thought the snake

might Tony laughed because he knew the

☐ 2

animal was a He said it was called

☐ 3

a

☐ 4

It went into the tent and crawled over Tony's airbed but when Tony touched his sleeping-bag, it ran quickly onto the next one. This one belonged

to It reared its head

☐ 5

and The baby yelled because he

☐ 6

was

☐ 7

The lizard then ran in to hide in the

☐ 8

Because Mother made a mistake about the lizard, she

felt and she was annoyed at

☐ 9

the that Tony was making.

☐ 10

Mother called out to Tony to look out for the eggs but it was too late because

Tony had already Tony

☐ 11

felt because he

☐ 12

had the lizard. His feet were goldy-yellow

☐ 13

because he had Tony had

☐ 14

to the eggs. He wished the lizard

☐ 15

was a because they love eggs and

☐ 16

would

☐ 17

He mopped up the mess and put it in the

☐ 18

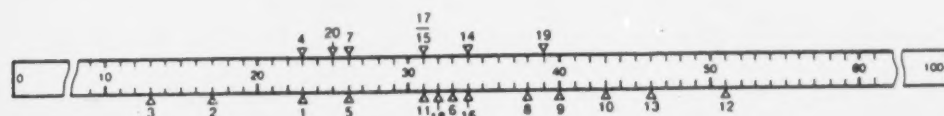
with the lizard. The lizard's tongue flicked in and out

as it all the broken eggs. So Tony found out

☐ 19

that it's not just goannas who like eggs, so do

☐ 20



BEST COPY AVAILABLE

60

☐ TOTAL

☐ TORCH SCORE

$\pm$ ☐ ERROR

Feeding Puff

The first cage in the row belonged to a couple of baby red river hogs which I had called Puff and Blow, and they were the most charming pair of babies imaginable. A full-grown red river hog is about the most colourful and handsome of the pig family. Its fur is a rich orange-red colour and along its back and neck is a mane of pure white fur; on the tips of its long, pointed ears are two dangling tufts of white hair. Puff and Blow, however, like all baby piglets, were striped; they were a dark chocolate brown, and their stripes were a light buttercup yellow, running from nose to tail. This made them look like fat little wasps, as they trotted round their pen.

Puff was the first one to arrive at the camp. He was brought in one morning, sitting rather sadly in a wicker basket balanced on the head of a native hunter. He had been captured in the forest, and I soon discovered the reason for his doleful appearance was that he had eaten nothing for two days, a thing that was enough to make any self-respecting pig look down in the snout. The hunter, who had caught him, had tried to feed him on bananas but Puff was far too young for that sort of food. What he wanted was milk, and plenty of it. So as soon as I had paid for him, I mixed a big bottleful of warm milk with sugar, and taking Puff on to my knees, I tried to make him drink. He was about the size of a Pekinese, with very small hooves and a pair of sharp little tusks as well, as I soon found out to my cost.

Of course, he had never seen a feeding bottle before, and treated it with the gravest suspicion from the start. When I lifted him on to my knees and tried to put the rubber into his mouth, he decided that this was some special kind of torture I had invented for him. He screamed and squealed, kicking me with his sharp little hooves and trying to stab me with his little tusks. After the struggle between us had lasted for about five minutes, both Puff and I looked as though we had been bathed in milk, but not a single drop of it had gone down his throat.

I filled another bottle and again grasped the squealing pig firmly between my knees, wedged his mouth open with one hand and started to squirt the milk in with the other. He was so busily squeaking for help that every time the milk was squirted into his mouth, the next squeal would spit it all out again. At last I was fortunate enough to get a few drops to trickle down his throat, and waited for him to get the taste of it, which he soon made apparent by stopping to yell and struggle, and by starting to smack his lips and grunt. I dribbled a little more milk into his mouth and he sucked it down greedily, and within a short while he was pulling away at the bottle as though he would never stop, while his tummy grew bigger and bigger. At length, when the last drop had disappeared from the bottle, he heaved a long sigh of satisfaction and fell into a deep sleep on my lap, snoring like a hive full of bees.

After that he was no more trouble, and after a few days had lost all his fear of humans, and would run, grunting and squeaking delightedly, to the bars of his pen when he saw me coming, and flop over on his back to have his tummy scratched. At feeding time, when he saw the bottle coming, he would push his nose through the bars and scream shrilly with excitement, and, to hear him, you would think he had never had a square meal in his life.



Feeding Puff

Helpful Hints

Here are some helpful hints for you to follow:

- Use a pencil and rubber so that you can change your answers if you want to.
- You can look back in the book as often as you like to find the answers.
- Often the words you want for your answer are not in the book, so feel free to use your own words.
- You can write one word for your answer or you can write as many words as you want in each space.
- If you can't fit all of your answer in the space ask for help.
- If you spell a word wrongly it doesn't mean the answer will be marked wrong. Ask for help with spelling if you really can't write the word you want.
- Don't get stuck on an answer you can't fill in. Leave it and come back to it later.
- When you have finished read through the exercise to check that it makes sense.

Practice Example

Septimus Potter lived all alone and that was how he liked it. Of course, there was his black cat Tobias; but Tobias was also a loner, and he was such a good hunter that Septimus didn't even have to feed him very often. Sometimes, when he found feathers scattered around the lily pond, Septimus cursed Tobias; but otherwise they lived in silence.

The passage is about a man called _____ who
He lived with his _____
sometimes upset his master because _____

NAME.....

YEAR.....

Feeding Puff

The camp had a number of cages where the animals were kept. In the first cage were two baby red river hogs called Puff and Blow. Their fur was striped in colours of and and they had small hooves, pointed ears and sharp When these animals grow to full size they have fur except on their backs where the fur is

Puff was captured and carried in a to the camp by The little animal was looking very sad when he arrived because he hadn't

The man knew that Puff had been given to eat but he hadn't eaten them because he was and could only have So he prepared a bottle for Puff. When the man held him Puff began to

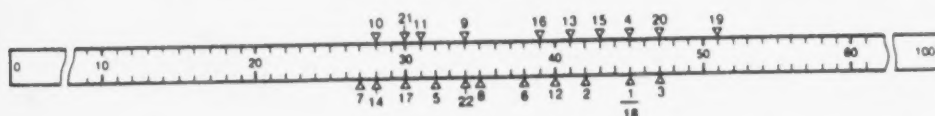
The first bottle was soon empty but the little hog was still struggling and hadn't The man filled another bottle. This time he managed to squirt milk into Puff's mouth but every time the little hog squealed, he Eventually some milk went

As soon as he , the little hog stopped and began to grunt and smack his lips. Soon he was drinking and his tummy began

Puff stopped causing trouble and after a few more days he was no longer Whenever the man walked by, Puff would run to the bars and so that the man would Puff became excited when he saw the man with because this meant it was

The little hog had settled in to living in a cage and being close to humans.

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10
- ☐ 11
- ☐ 12
- ☐ 13
- ☐ 14
- ☐ 15
- ☐ 16
- ☐ 17
- ☐ 18
- ☐ 19
- ☐ 20
- ☐ 21
- ☐ 22



☐ TOTAL
☐ TORCH SCORE
 ± ☐ ERROR

The Cats

Jim heard a sudden hiss behind him, and almost simultaneously Socker leaned over and clapped the palm of his hand over Kevin's mouth. He used his favourite phrase. 'Shut up,' he whispered in Kevin's ear.

For a long period, it seemed, they watched the cat and the cat watched them. Sometimes it licked its lips. Sometimes it blinked. Once it got up, stretched, and sat down again. But it watched them all the time. For their part they found it impossible to take their eyes from it. Then Jim got a drop of water in his eye and rubbed it. He rolled both eyes to clear his vision, and his cleared vision showed him another cat. It, also, was in full view and it was lying on the path behind them on the edge of the mist. This one was a grey cat and it melted into the general greyness. It lay with its head on its paws. He might have thought it asleep if its yellow eyes had not been wide open looking at them.

'There's a cat behind us,' he whispered to Socker.

Socker had nothing to say, but his face became more tight and strained. Afterwards one of the things Jim was able to remember was the way the raindrops clung like seed pearls to Socker's strong dark hair.

Not long afterwards they picked out two more cats — tabbies — side by side on the other side of the path. Their eyes, too, were fixed on them. Little by little they picked out more and more cats through the shrouding mist. They never saw them come. They never heard them. But, suddenly, there they were. The three huddled figures pressed closer together now, waiting as the cats seemed to be waiting. The only difference was that the cats knew what they were waiting for.

Once Kevin whispered, 'They can't be cats. They're too big.' Socker whispered back, 'Who cares what you call them? They're there, aren't they?'

Jim wondered what would happen if he simply got up and walked on. Would they let him through? Would they let them all through? He did not think the other two would follow him and when he thought of himself alone in the mist among the cats he knew he could not do it. In any case, going on achieved nothing. He wished Willy would come. He never doubted Willy would come at last.

The mist hung about them, impenetrable, wet and silent. And the cats waited and watched, silent too. There were perhaps ten or twelve visible now and on the damp air came the smell of them; not of tomcat, but the nerve-tingling feral smell of zoo. Every now and then one would get up and shake itself free of raindrops and lie down again, and each time, when they saw the size of the standing cat they huddled a little closer together. On the other side Kevin pressed even closer, and he was cold and beginning to tremble. The mist began to get thicker. It hung dark and oppressive overhead and pushed in on all sides. It was harder to see the cats now, but the cats apparently had no trouble seeing them, for they moved no closer. A kind of bubble burst inside Jim when he realized the mist was getting no thicker. It was simply that the light was fading. It was getting dark. The night was coming and soon they would be able to see nothing at all. Then they would be helpless. But the cats — the cats could see in the dark. The lack of light made no difference to them.

It was when it had become too dark to see the cats' eyes that Kevin lost his nerve. They had been sitting silent and still for so long, the target of so many yellow eyes, conscious that each tiny movement any of them made was seen and noted. Kevin had begun to whimper softly, saying his bitten hand was throbbing and that he felt ill. As usual Socker said, 'Shut up.' This time it had the opposite effect. Kevin jumped up, began to make high, incomprehensible sounds and started to run down the road. In a moment he had disappeared into the mist, but they could hear his feet, still running. At the same time, as if this was what they

had been waiting for, the cats came to life. They were all on their feet before Kevin had been swallowed by the mist, and Jim saw several of them bound down the road after him.

Jim shouted, 'Kevin, stop!' but his voice hardly carried at all through the mist. There were still cats surrounding them, and these began to move now, closing in and making a high, wavering singing sound, eerie and sinister in the deepening night. There was no chance of following Kevin if they had wanted to. At any moment they would be doing battle on their own account. Then, just as Kevin's voice reached a higher, shriller note, something rushed out of the mist on their other side. For a moment, pressed as he was, his senses strained to their limit, Jim thought the figure that sprang towards them was taller than any normal person had a right to be, and that there was a flickering light about it. Then he blinked and looked again, and it was only Willy, running towards them and shouting in a commanding voice. As he reached them the cats drew back and their singing stopped.

'Come with me,' said Willy. 'We'd better get to Kevin.' He ran on, and Jim and Socker followed.

They ran on to Kevin almost before they saw him. He was still screaming and there were cats all round him. One had his injured hand in its mouth and another had him by the ankle. So far they had done no more. Now, as Willy arrived, they let go and drew back. But they stood, watching, tails twitching, while Kevin sank to the ground and moaned.

'I knew you'd come,' said Jim.

'Get up,' Willy said to Kevin. 'You're not hurt.' Kevin stopped moaning and got up, nursing his hand. 'It wasn't you they're after. But you shouldn't have run. Don't you know a cat will always chase anything that runs?'

'What are they after, then?' said Socker.

Willy turned and faced him. He still looked tall in the deceptive half-light. 'I think they're after you, Socker,' he said.



The Cats

Helpful Hints

Here are some helpful hints for you to follow:

- Use a pencil and rubber so that you can change your answers if you want to.
- You can look back in the book as often as you like to find the answers.
- Often the words you want for your answer are not in the book, so feel free to use your own words.
- You can write one word for your answer or you can write as many words as you want in each space.
- If you can't fit all of your answer in the space ask for help.
- If you spell a word wrongly it doesn't mean the answer will be marked wrong. Ask for help with spelling if you really can't write the word you want.
- Don't get stuck on an answer you can't fill in. Leave it and come back to it later.
- When you have finished read through the exercise to check that it makes sense.

Practice Example

Septimus Potter lived all alone and that was how he liked it. Of course, there was his black cat Tobias; but Tobias was also a loner, and he was such a good hunter that Septimus didn't even have to feed him very often. Sometimes, when he found feathers scattered around the lily pond, Septimus cursed Tobias; but otherwise they lived in silence.

The passage is about a man called _____
He lived with his _____ who
sometimes upset his master because _____

NAME.....

YEAR.....

The Cats

The situation that the writer is describing in this story is quite strange. The boys were very of the cats and they were having trouble seeing clearly because They had surrounded the and were standing and staring at them. Kevin and Jim were together. They were waiting for the other boy named who Jim seemed to think could them.

- ☐ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ 6

When..... gave into his and ran off, some of the cats and the other cats Jim and Socker. The boys were too to run after until arrived. When the boys did go to Kevin, the cats were him but at the sight of they Kevin. This may have been because the cats

- ☐ 7, 8
☐ 9
☐ 10
☐ 11, 1
☐ 13
☐ 14
☐ 15, 1
☐ 17

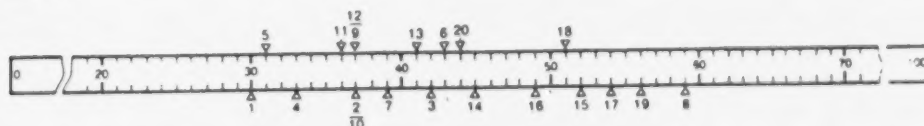
Willy told Kevin that he shouldn't have run because even though the cats chased him they weren't

☐ 18

Willy thought the cats were really and that they had only chased Kevin because cats

- ☐ 19
☐ 20

The boys were finally together but they still had to deal with the problem of the cats.



- ☐ TOTAL
☐ TORCH SCO
 $\pm$ ☐ ERROR

BEST COPY AVAILABLE

Appendix C: Student ability compared with the difficulty of passages used in the TORCH reading comprehension test

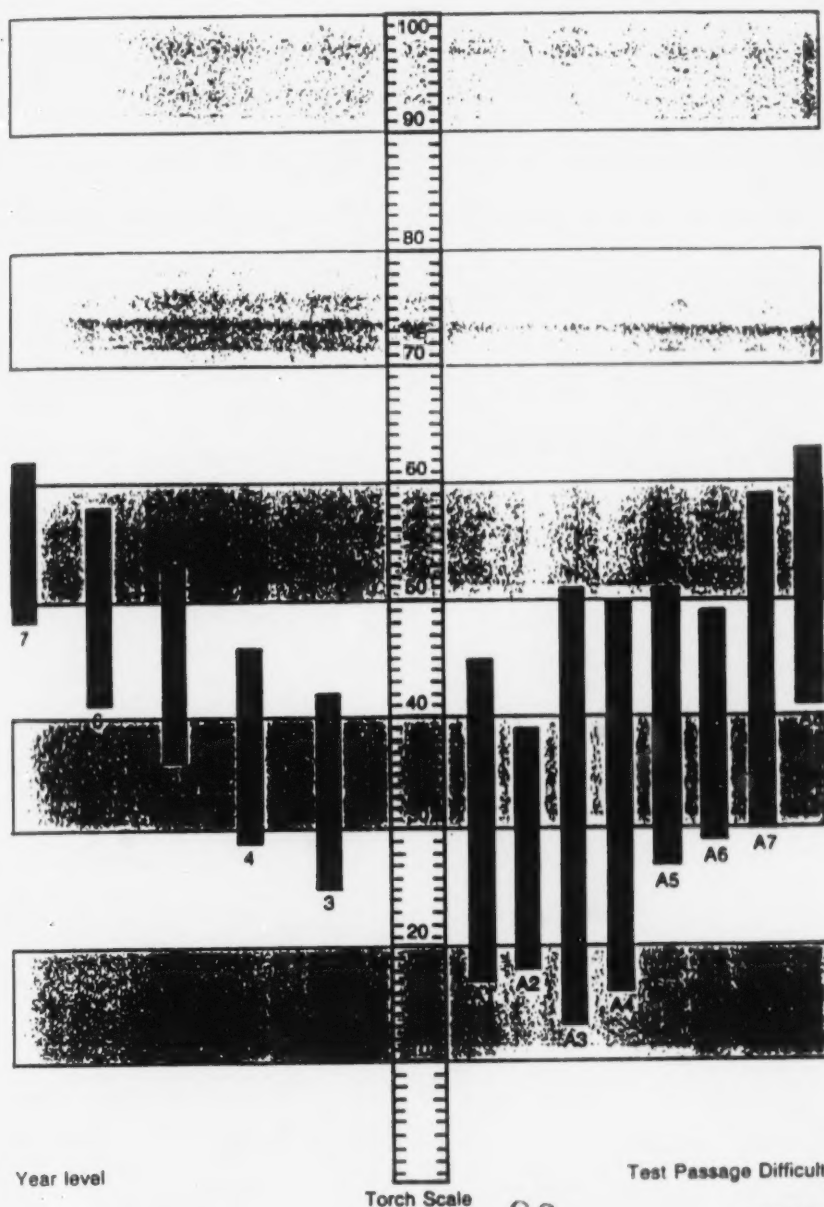
This appendix shows the ranges of reading ability associated with each passage of text used by TORCH.

(This information is taken directly from the TORCH manual.) The relevant passages are:

A3: *Lizards love eggs*

A5: *Feeding Puff*

A7: *The Cats*



Appendix D: Writing exercise materials

This appendix contains all the materials sent to classroom teachers for the administration of the writing exercise. This appendix includes the following:

- Teachers information and instruction sheet for the administration of the writing exercise
- Student report form for teachers to list the names of students with special circumstances which may have affected their performance on the test
- Writing exercise sheet for students
- Sheet for students to write redraft of writing exercise.

Information for teachers about the Writing Exercise

There are two stages to this writing exercise. The first stage consists of the children writing a draft report. This stage has a maximum time limit of 30 minutes. The second stage consists of the students redrafting their answer using the normal methods of this classroom for redrafting. This second stage is not to be timed unless timing forms part of the normal redrafting procedures students use in this classroom.

Stage 1: The draft report

1. Distribute the writing exercise sheets to each Year 5 student in this class.
2. Ensure that each Year 5 student has a pencil, biro or pen, and an eraser. Spare pencils and erasers should be available if possible.
3. Say to the students;

"I am going to ask you to write a draft report about an animal. You will have 30 minutes to complete your draft. Do not use other materials like dictionaries to help you at this time. You will get a chance later to redraft what you write today."
4. Ask for and then answer any questions.
5. Instruct the students to write their name and school on the writing exercise sheet. When they have done this, have them start.
6. When the thirty minutes are up ask the students to stop work.
10. Collect the papers and check that each student has put his or her name and school on the front page.

Stage 2: Redrafting the report

It is expected that students will find the writing task quite demanding. It is probably better, therefore, to wait for another day before asking them to redraft it. However, if it is normal procedure in the class for the redrafting to be done soon after the first draft then this should be done.

In either case, on another day or at another time, return the draft report to the students and have them redraft their report. This redraft should be done on the sheet of paper provided for this purpose. There is no time limit on the redraft and students should use any of the normal methods of this class to prepare their redraft. If this redrafting process involves marking the original version, ensure that a different coloured pen or pencil is used so that the original can be unambiguously identified.

1. Return writing task sheets to students.
2. Distribute Redraft sheets to students.
3. Ask students to write their name and school on the Redraft sheet.
4. Ask students to redraft their report. They may have access to any materials or resources that they would normally use.
5. When the redrafts are finished have each student return both the draft and the final version to you.
6. Complete the Student Report describing which students may have special circumstances that could effect their performance on the writing task.
7. Bundle up (1) the completed drafts and redrafts, and (2) the completed student report and, using the envelope supplied return to:

Adrian Harvey-Beavis
First Steps Evaluation
Reply Post 2
Australian Council for Education Research
PO Box 210
Hawthorn
Victoria 3122

If this address is used then no postage stamp will be required.

The results of the writing exercise will be returned to you as soon as possible.

Thank you for your help.

**First Steps Evaluation, 1992:
Writing Exercise Report Form**

Name of School:

Please note here the name of any student with special circumstances that may have affected his/her performance in completing the writing exercises. In particular indicate students who have a physical disability, limited command of the English language, etc.

Name of Student

Special circumstances

.....	
.....	
.....	
.....	
.....	
.....	
.....	
.....	
.....	
.....	
.....	
.....	

Briefly describe the methods the Year 5 students use in redrafting their written reports.

Thank you.

Writing Exercise

SCHOOL:

73

Published by
The Australian Council for Educational Research Limited
Radford House, Frederick Street, Hawthorn, Victoria 3122
Copyright 1922 ACER.

Redraft of Report on an Animal

Name:

School:

Use the space below to write a new draft of your answer to the report you wrote about an animal of your choice..

Appendix E: The ACER criteria for assessment of written expression

In assessing written expression markers at ACER will consider the following issues:

Thought and content:

(the quality of what is said in the piece of writing)

- what is made of and developed from the task
- the kinds of thoughts and feelings offered in response to the task

Structure and organisation:

(the quality of the structure and organisation developed to present what is said in the writing)

- the shape and form of the piece
- the sequence and cohesion of the piece

Expression, style and mechanics

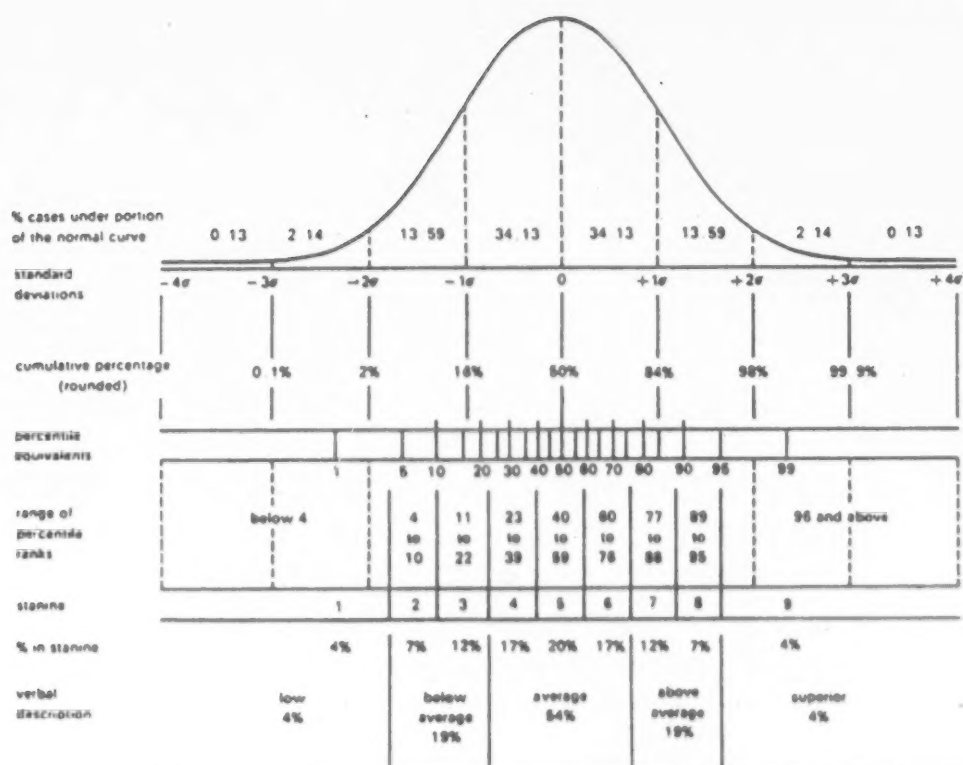
(the quality of the language used to organise and present what is said)

- the effectiveness and appropriateness of the language
- the expressiveness and fluency of the language
- the control of the mechanics of English

Appendix F: Estimated percentile ranks of TORCH reading comprehension scores

The table and figure in this appendix are taken from the TORCH manual.

TORCH score	Estimated Percentile Ranks							
	Year							
	3	4	5	6	7	8	9	10
95								
90								
85								
80						98		93
75					98	94	93	82
70					93	83	82	67
65			98	93	84	71	64	44
60		98	91	83	68	51	45	28
55	97	94	80	61	46	34	29	14
50	92	84	65	48	27	19	16	7
45	85	72	47	33	14	11	8	4
40	70	55	34	20	7	6	5	
35	56	40	20	10	3	3	3	
30	40	25	11	6				
25	24	15	6	3				
20	15	7	2					
15	9	2						
10	5							
5								
Mean	33.2	37.9	44.8	50.0	55.5	59.0	60.3	65.5
SD	12.5	11.6	11.6	11.2	9.9	11.2	10.9	10.2
N	773	903	738	831	681	665	682	635



Appendix G: The reasons for (and the implications of) not using tests of statistical significance in this report

This appendix explains why tests of statistical significance are not used in this report. It also briefly describes the implications of not using these tests.

In explaining why tests of statistical significance are not used with the data, reported here, the question really being addressed is: Why is the ANOVA procedure not being used with these data? An ANOVA is the statistical procedure which would be expected to be applied to these data. It provides a measure of the probability that the observed differences between two or more mean scores occurred by chance. The approach taken in this report to the analysis of the data parallels that taken when using an ANOVA statistical test. In other words, it is concerned with identifying what are, or might be, the main effects and the interaction effect in the data. But it does not use the ANOVA procedure.

ANOVA cannot be used here because there was no random assignment of students to the various levels of involvement with First Steps. First Steps personnel advise that those most disadvantaged were selected to go into the program first (the students in old First Steps schools), and those least disadvantaged were not put into the program. This lack of randomness would not be such a problem if the TORCH reading scores of the students were known before First Steps was implemented. However, without knowing the scores of students before the introduction of First Steps it is not possible to generate an assumption to test using the ANOVA procedure. For example, in the orthodox experimental case, it is assumed that the various groups that are being tested come from the same population and so will have about the same mean score and variance. The assumption to be tested in this case is that there is no difference between the various groups. If a statistically significant difference is found, then this assumption can be rejected. In rejecting this assumption one can tentatively accept the alternative to the assumption of no difference and claim that there probably is a difference between the groups. But it is not necessary to assume that the differences between the groups are equal to zero. Any number of differences could be assumed. For example, it could be assumed that because students in old First Steps schools were the most disadvantaged, they would have lower average reading ability than students from non First Steps schools who had highest average reading ability. However, without knowing what these differences are it is not possible to establish precisely what assumption to make in order to test it. (The differences need to be specified as a set of specific values.) Consequently, as these values are unknown for the data being examined here, it is not possible to use ANOVA and have the results be meaningful. (In fact in the report some possible values are guessed at and the implication of these (informed) guesses are examined. It was judged to be inappropriate to apply statistical tests to these guesses because this could bestow an unwarranted aura of scientificity.) Because it is not possible to use the ANOVA

procedure it is also not possible to establish the statistical significance of the differences between the means. Following directly from this, it is not possible to establish whether the differences observed in the sample of data are likely to arise from random variation in the TORCH scores or from real differences between the groups.

Appendix H: A description of the ANOVA procedure

In using ANOVA one usually starts with the assumption that there is no real difference between the mean scores of the groups. (This is the same as starting with the assumption that the different groups being examined come from the same population and that the observed differences between groups are a function of random variation within the population.) If it is shown that the observed differences between the groups are too large to result from random variation then the assumption that the two groups come from the same population is rejected. The ANOVA procedure is used to establish how likely it is that the differences are due to random variation.

Main effects and interaction effects in the data

When an ANOVA is performed there are, usually, two types of 'effects' which are examined. These are called 'main effects' and 'interaction effects'. An interaction effect is so named because it arises from the interaction between some or all of the main effects. Each of these types of effects is now described.

For each independent variable¹ or factor under consideration there can be a main effect. With the TORCH reading ability data used here there are two factors and so two main effects. These factors are the socio-economic status background of the students and the level of involvement of the students with First Steps.² Each factor has a number of different levels. The 'socio-economic status background' factor has two levels - 'low' and 'middle or high'. The 'type of involvement with First Steps' factor has three levels. These are (1) no involvement (children at non First Steps schools), (2) less than one year's involvement (children at new First Steps schools) and (3) more than one year's involvement (children at old First Steps schools).

The main effect of a factor is seen in the differences between the mean score of each level of that factor.³ For example, if the mean score of all students with a middle or high socio-economic status background in the sample is compared with the mean score of all students with a low socio-economic status background then the main effect of socio-economic status background can be established. With the data reported here, the mean TORCH score for the

¹ An independent variable is a variable which, in an experiment is the 'treatment' administered to a subject. The experiment is designed to establish the effect of this independent variable.

² Clearly, socio-economic status background is not a 'treatment' in the classical experimental tradition. Nevertheless, it is regarded as such here. It is probably more obvious that what is also being examined here is the effect of different amounts of treatment using First Steps.

³ More precisely, the effect of a factor is seen when the differences between the means are examined while taking into account the variance in the scores. Generally, the greater the variance the bigger the difference has to be between the means before they will be defined as statistically significant. If a difference is statistically significant then it is usually asserted that there is probably an effect in the data. If the difference is not statistically significant, then there is said to be probably no effect in the data.

middle or high socio-economic status background group is 46.7 and the mean score for the low socio-economic status background group is 43.4. This difference between the means can be interpreted as follows. Provisionally,⁴ the effect of socio-economic status background is to move the mean TORCH score of students with middle or low socio-economic status background above the grand mean (46.0) and to move the mean TORCH score of students from a low socio-economic status background below the grand mean. If only data about socio-economic status background and TORCH scores were collected and if these differences could be shown to be statistically significantly different, then it could be claimed that socio-economic status background seems to affect reading comprehension scores.

An interaction effect occurs when two (or more) variables or factors interact such that their combined effect is different from their overall effects. Consider a simple example.

Assume that the following (somewhat depressing) hypothetical data were collected:

Mean number of deaths per 1000 adults per year caused by smoking	= 10
Mean number of deaths per 1000 adults per year caused by alcohol abuse	= 10

What would be expected of that group of persons whom both smoked and abused alcohol? It might be expected that, on average, there would at the end of a year be a set of 10 dead smokers (who happened to abuse alcohol) and a set of 10 dead alcohol abusers (who happened to smoke) from a group of 1000 smoking alcohol abusers. It is possible, but unlikely, that a member of the set of dead smokers was also a member of the set of dead drinkers. As a result, it could be expected that there would be about 20 deaths per 1000 per year from the population of smoking alcohol abusers related to these factors. However, if the data showed that this group of smoking alcohol abusers died at an average rate of 30 per 1000 persons per year then it could be concluded that smoking and alcohol are probably amplifying the effects of each other. In circumstances such as this, there is said to be an interaction. The importance of an interaction effect is that it provides a clearer or more precise picture of the world. For this reason, if an interaction effect can be demonstrated, then it takes precedence over the main effects in any explanations of the data. Consequently, any discussion of a main effect has to be qualified by the interaction effects. So, for example, if it was claimed that persons who smoked died at an average rate of 10 per 1000 persons per year, once the interaction effect with alcohol abuse was known, for this claim to still be true it would have to be qualified by the statement that it applied only to non alcohol abusing smokers. The interaction effect, once established, allows the claim that, while smoking causes 10 deaths per 1000 per year, alcohol abuse worsens its

⁴It is provisional until a test of significance can establish the probability of the observed difference between means occurring by chance. Even if a statistically significant difference is found, it is still possible that the observed difference will not be found in the population but has occurred because of a remote chance.

impact. On these hypothetical data, if a person abuses alcohol and smokes, they are three times more likely to die in the following year than if they just smoked or if they just abused alcohol. They are one and a half times more likely to die than if there were no interaction effect.

An educational example now, which is more complex.. This example uses mean scores.

Imagine that the data described in Exhibit 1H were collected measuring the reading ability of Year 3 and Year 5 male and female students.

Exhibit 1H: Hypothetical data showing the mean reading ability scores of students in Year 3 and Year 5 for each sex.

	Year 3	Year 5	
Male	10	18	14
Female	12	14	13
	11	16	

With these data, the main effects show that, overall, males (14) have a higher reading ability score than females (13) and that Year 5 (16) have a higher reading ability than Year 3 (11). However, an examination of the body of the table shows that it is only in Year 5 that males (18) score higher than females (14). In Year 3, it is females (12) who score higher than males (10). These data thus show that there is an interaction between sex and Year level. The generalisation that males score more highly than females has to be qualified now by the additional information that this claim is dependent upon the Year level of the students.